RATE OF CONVERGENCE AND EDGEWORTH-TYPE EXPANSION IN THE ENTROPIC CENTRAL LIMIT THEOREM

S. G. BOBKOV[1,4], G. P. CHISTYAKOV[2,4], AND F. GÖTZE[3,4]

Abstract. An Edgeworth-type expansion is established for the entropy distance to the class of normal distributions of sums of i.i.d. random variables or vectors, satisfying minimal moment conditions.

1. Introduction

Let $(X_n)_{n\geq 1}$ be independent identically distributed random variables with mean $\mathbf{E}X_1 = 0$ and variance $\mathrm{Var}(X_1) = 1$. According to the central limit theorem, the normalized sums

$$Z_n = \frac{X_1 + \cdots + X_n}{\sqrt{n}}$$

are weakly convergent in distribution to the standard normal law: $Z_n \Rightarrow Z$, where $Z \sim N(0,1)$ with density $\varphi(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}$. A much stronger statement (when applicable) – the entropic central limit theorem – indicates that, if for some n_0, or equivalently, for all $n \geq n_0$, Z_n have absolutely continuous distributions with finite entropies $h(Z_n)$, then there is convergence of the entropies,

$$h(Z_n) \rightarrow h(Z), \quad \text{as} \ \ n \rightarrow \infty. \tag{1.1}$$

This theorem is due to Barron [Ba]. Some weaker variants of the theorem in case of regularized distributions were known before; they go back to the work of Linnik [L], initiating an information-theoretic approach to the central limit theorem.

To clarify in which sense (1.1) is strong, first let us recall that, if a random variable X with finite second moment has a density $p(x)$, its entropy

$$h(X) = -\int_{-\infty}^{+\infty} p(x) \log p(x)\, dx$$

is well-defined and is bounded from above by the entropy of the normal random variable Z, having the same mean a and the variance σ^2 as X (the value $h(X) = -\infty$ is possible). The

1991 *Mathematics Subject Classification.* Primary 60E.
Key words and phrases. Entropy, entropic distance, central limit theorem, Edgeworth-type expansions.
1) School of Mathematics, University of Minnesota, USA; Email: bobkov@math.umn.edu.
2) Faculty of Mathematics, University of Bielefeld, Germany; Email: chistyak@math.uni-bielefeld.de.
3) Faculty of Mathematics, University of Bielefeld, Germany; Email: goetze@math.uni-bielefeld.de.
4) Research partially supported by SFB 701.

relative entropy

$$D(X) = D(X||Z) = h(Z) - h(X) = \int_{-\infty}^{+\infty} p(x) \log \frac{p(x)}{\varphi_{a,\sigma}(x)}\,dx,$$

where $\varphi_{a,\sigma}$ stands for the density of Z, is non-negative and serves as kind of a distance to the class of normal laws, or to Gaussianity. This quantity does not depend on the mean or the variance of X, and may be related to the total variation distance between the distributions of X and Z by virtue of the Pinsker-type inequality $D(X) \geq \frac{1}{2}\,\|F_X - F_Z\|_{\rm TV}^2$. This already shows that the entropic convergence (1.1) is stronger than in the total variation norm.

Thus, the entropic central limit theorem may be reformulated as $D(Z_n) \rightarrow 0$, as long as $D(Z_{n_0}) < +\infty$ for some n_0. This property itself gives rise to a number of intriguing questions, such as to the type and the rate of convergence. In particular, it has been proved only recently that the sequence $h(Z_n)$ is non-decreasing, so that $D(Z_n) \downarrow 0$, cf. [A-B-B-N1], [B-M]. This leads to the question as to the precise rate of $D(Z_n)$ tending to zero; however, not much seems to be known about this problem. The best results in this direction are due to Artstein, Ball, Barthe and Naor [A-B-B-N2], and to Barron and Johnson [B-J]. In the i.i.d. case as above, they have obtained an expected asymptotic bound $D(Z_n) = O(1/n)$ under the hypothesis that the distribution of X_1 admits an analytic inequality of Poincaré-type (in [B-J], a restricted Poincaré inequality is used). These inequalities involve a large variety of "nice" probability distributions which necessarily have a finite exponential moment.

The aim of this paper is to study the rate of $D(Z_n)$, using moment conditions $\mathbf{E}\,|X_1|^s < +\infty$ with fixed values $s \geq 2$, which are comparable to those required for classical Edgeworth-type approximations in the Kolmogorov distance. The cumulants

$$\gamma_r = i^{-r}\,\frac{d^r}{dt^r}\,\log \mathbf{E}\,e^{itX_1}|_{t=0}$$

are then well-defined for all $r \leq [s]$ (the integer part of s), and one may introduce the functions

$$q_k(x) \ = \ \varphi(x) \sum H_{k+2j}(x)\,\frac{1}{r_1! \dots r_k!}\,\Big(\frac{\gamma_3}{3!}\Big)^{r_1} \dots \Big(\frac{\gamma_{k+2}}{(k+2)!}\Big)^{r_k} \tag{1.2}$$

involving the Chebyshev-Hermite polynomials H_k. The summation in (1.2) runs over all non-negative solutions $(r_1,\dots,r_k)$ to the equation $r_1 + 2r_2 + \dots + kr_k = k$ with $j = r_1 + \dots + r_k$.

The functions q_k are defined for $k = 1,\dots,[s]-2$. They appear in the Edgeworth-type expansions including the local limit theorem, where q_k are used to construct the approximation of the densities of Z_n. These results can be applied to obtain an expansion in powers of $1/n$ for the distance $D(Z_n)$. For a multidimensional version of the following Theorem 1.1 for moments of integer order $s \geq 2$, see Theorem 6.1 below.

Theorem 1.1. *Let $\mathbf{E}\,|X_1|^s < +\infty$ $(s \geq 2)$, and assume $D(Z_{n_0}) < +\infty$, for some n_0. Then*

$$D(Z_n) = \frac{c_1}{n} + \frac{c_2}{n^2} + \dots + \frac{c_{[(s-2)/2)]}}{n^{[(s-2)/2)]}} + o\big((n \log n)^{-(s-2)/2}\big). \tag{1.3}$$

Here

$$c_j = \sum_{k=2}^{2j} \frac{(-1)^k}{k(k-1)} \sum \int_{-\infty}^{+\infty} q_{r_1}(x) \dots q_{r_k}(x)\,\frac{dx}{\varphi(x)^{k-1}}, \tag{1.4}$$

where the summation runs over all positive integers $(r_1,\dots,r_k)$ such that $r_1 + \dots + r_k = 2j$.

The implied error term in (1.3) holds uniformly in the class of all distributions of X_1 with prescribed tail behaviour $t \to \mathbf{E}\,|X_1|^s\, 1_{\{|X_1|>t\}}$ and prescribed values n_0 and the entropy distance $D(Z_{n_0})$.

As for the coefficients, each c_j represents a certain polynomial in the cumulants $\gamma_3, \dots, \gamma_{2j+1}$. For example, $c_1 = \frac{1}{12}\,\gamma_3^2$, and in the case $s = 4$ (1.3) gives

$$D(Z_n) = \frac{1}{12\,n}\,\big(\mathbf{E}X_1^3\big)^2 + o\left(\frac{1}{n\log n}\right) \qquad \big(\mathbf{E}X_1^4 < +\infty\big). \tag{1.5}$$

Thus, under the 4-th moment condition, as conjectured by Johnson ([J], p. 49), we have $D(Z_n) \leq \frac{C}{n}$, where the constant depends on the underlying distribution. Actually, this constant may be expressed in terms of $\mathbf{E}X_1^4$ and $D(X_1)$, only.

When s varies in the range $4 \leq s \leq 6$, the leading linear term in (1.5) will be unchanged, while the remainder term improves and satisfies $O(\frac{1}{n^2})$ in case $\mathbf{E}X_1^6 < +\infty$. But for $s = 6$, the result involves the subsequent coefficient c_2 which depends on γ_3, γ_4, and γ_5. In particular, if $\gamma_3 = 0$, we have $c_2 = \frac{1}{48}\,\gamma_4^2$, thus

$$D(Z_n) = \frac{1}{48\,n^2}\,\big(\mathbf{E}X_1^4 - 3\big)^2 + o\left(\frac{1}{(n\log n)^2}\right) \qquad \big(\mathbf{E}X_1^3 = 0,\ \mathbf{E}X_1^6 < +\infty\big).$$

More generally, the representation (1.3) simplifies if the first $k-1$ moments of X_1 coincide with the corresponding moments of $Z \sim N(0,1)$.

Corollary 1.2. *Let $\mathbf{E}|X_1|^s < +\infty$ $(s \geq 4)$, and assume that $D(Z_{n_0}) < +\infty$, for some n_0. Given $k = 3, 4, \dots, [s]$, assume that $\gamma_j = 0$ for all $3 \leq j < k$. Then*

$$D(Z_n) \;=\; \frac{\gamma_k^2}{2k!}\cdot\frac{1}{n^{k-2}} + O\bigg(\frac{1}{n^{k-1}}\bigg) + o\bigg(\frac{1}{(n\log n)^{(s-2)/2}}\bigg). \tag{1.6}$$

Johnson has noticed (though in terms of the standardized Fisher information, see [J], Lemma 2.12) that if $\gamma_k \neq 0$, $D(Z_n)$ cannot be better than $n^{-(k-2)}$.

Note that when $\mathbf{E}X_1^{2k} < +\infty$, the o-term may be removed from the representation (1.6). On the other hand, when $k > \frac{s+2}{2}$, the o-term will dominate the $n^{-(k-2)}$-term, and we can only say that $D(Z_n) = o\big((n\log n)^{-(s-2)/2}\big)$.

As for the missing range $2 \leq s < 4$, there are no coefficients c_j in the sum (1.3), and Theorem 1.1 just tells us that

$$D(Z_n) = o\left(\frac{1}{(n\log n)^{(s-2)/2}}\right). \tag{1.7}$$

This bound is worse than the rate $1/n$. In particular, it only gives $D(Z_n) = o(1)$ for $s = 2$, which is the statement of Barron's theorem. In fact, in this case the entropic distance to normality may decay to zero at an arbitrarily slow rate. In case of a finite 3-rd absolute moment, $D(Z_n) = o(\frac{1}{\sqrt{n\log n}})$. To see that this and that the more general relation (1.7) cannot be improved with respect to the powers of $1/n$, we prove:

Theorem 1.3. *Let $\eta > 1$. Given $2 < s < 4$, there exists a sequence of independent identically distributed random variables $(X_n)_{n\geq 1}$ with $\mathbf{E}\,|X_1|^s < +\infty$, such that $D(X_1) < +\infty$*

and

$$D(Z_n) \geq \frac{c(\eta)}{(n \log n)^{(s-2)/2} \, (\log n)^{\eta}}, \qquad n \geq n_1(X_1),$$

with a constant $c(\eta) > 0$*, depending on* η*, only.*

Known bounds on the entropy and Fisher information are commonly based on Bruijn's identity which may be used to represent the entropic distance to normality as an integral of the Fisher information for regularized distributions (cf. [Ba]). However, it is not clear how to reach exact asymptotics with this approach. The proofs of Theorems 1.1 and 1.3 stated above rely upon classical tools and results in the theory of summation of independent summands including Edgeworth-type expansions for convolution of densities formulated as local limit theorems with non-uniform remainder bounds. For non-integer values of s, the authors had to complete the otherwise extensive literature by recent technically rather involved results based on fractional differential calculus, see [B-C-G]. Our approach applies to random variables in higher dimension as well and to non-identical distributions for summands with uniformly bounded s-th moments.

We start with the description of a truncation-of-density argument, which allows us to reduce many questions about bounding the entropic distance to the case of bounded densities (Section 2). In Section 3 we discuss known results about Edgeworth-type expansions that will be used in the proof of Theorem 1.1. Main steps of the proofs are based on it in Sections 4-5. All auxiliary results also cover the scheme of i.i.d. random vectors in $\mathbf{R}^d$ (however, with integer values of s) and are finalized in Section 6 to obtain multidimensional variants of Theorem 1.1 and Corollary 1.2. Sections 7-8 are devoted to lower bounds on the entropic distance to normality for a special class of probability distributions on the real line, that are applied in the proof Theorem 1.3.

2. **Binomial decomposition of convolutions**

First let us comment on the assumptions in Theorem 1.1. It may occur that X_1 has a singular distribution, but the distribution of $X_1 + X_2$ and of all next sums $S_n = X_1 + \dots + X_n$ ($n \geq 2$) are absolutely continuous (cf. [T]).

If it exists, the density p of X_1 may or may not be bounded. In the first case, all the entropies $h(S_n)$ are finite. If p is unbounded, it may happen that all $h(S_n)$ are infinite, even if p is compactly supported. But it may also happen that $h(S_n)$ is finite for some $n = n_0$ and then entropies are finite for all $n \geq n_0$ (see [Ba] for specific examples).

Denote by $p_n(x)$ the density of $Z_n = S_n/\sqrt{n}$ (when it exists). Since it is desirable to work with bounded densities, we will slightly modify p_n at the expense of a small change in the entropy. Variants of the next construction are well-known; see e.g. [S-M], [I-L], where the central limit theorem was studied with respect to the total variation distance. Without any extra efforts, we may assume that X_n take values in $\mathbf{R}^d$ which we equip with the usual inner product $\langle \cdot, \cdot \rangle$ and the Euclidean norm $|\cdot|$. For simplicity, we describe the construction in the situation, where X_1 has a density $p(x)$ (cf. Remark 2.5 on the appropriate modifications in the general case).

Let $m_0 \geq 0$ be a fixed integer. (For the purposes of Theorem 1.1, one may take $m_0 = [s]+1$.)

If p is bounded, we put $\widetilde{p}_n(x) = p_n(x)$ for all $n \geq 1$. Otherwise, the integral

$$b = \int_{p(x)>M} p(x)\,dx \tag{2.1}$$

is positive for all $M > 0$. Choose M to be sufficiently large to satisfy, e.g., $0 < b < \frac{1}{2}$ (cf. Remark 2.4). In this case (when p is unbounded), consider the decomposition

$$p(x) = (1-b)\rho_1(x) + b\rho_2(x), \tag{2.2}$$

where ρ_1, ρ_2 are the normalized restrictions of p to the sets $\{p(x) \leq M\}$ and $\{p(x) > M\}$, respectively. Hence, for the convolutions we have a binomial decomposition

$$p^{*n} = \sum_{k=0}^{n} C_n^k\,(1-b)^k\,b^{n-k}\,\rho_1^{*k} * \rho_2^{*(n-k)}.$$

For $n \geq m_0 + 1$, we split the above sum into the two parts, so that $p^{*n} = \rho_{n1} + \rho_{n2}$ with

$$\rho_{n1} = \sum_{k=m_0+1}^{n} C_n^k\,(1-b)^k\,b^{n-k}\,\rho_1^{*k} * \rho_2^{*(n-k)}, \quad \rho_{n2} = \sum_{k=0}^{m_0} C_n^k\,(1-b)^k\,b^{n-k}\,\rho_1^{*k} * \rho_2^{*(n-k)}.$$

Note that, whenever $b < b_1 < \frac{1}{2}$,

$$\varepsilon_n \equiv \int \rho_{n2}(x)\,dx = \sum_{k=0}^{m_0} C_n^k\,(1-b)^k\,b^{n-k} \leq n^{m_0}\,b^{n-m_0} = o(b_1^n), \qquad \text{as} \ \ n \to \infty. \tag{2.3}$$

Finally define

$$\widetilde{p}_n(x) = p_{n1}(x) = \frac{1}{1-\varepsilon_n}\,n^{d/2}\,\rho_{n1}\big(x\sqrt{n}\big), \tag{2.4}$$

and similarly $p_{n2}(x) = \frac{1}{\varepsilon_n}\,n^{d/2}\,\rho_{n2}(x\sqrt{n})$. Thus, we have the desired decomposition

$$p_n(x) = (1-\varepsilon_n)p_{n1}(x) + \varepsilon_n p_{n2}(x). \tag{2.5}$$

The probability densities $p_{n1}(x)$ are bounded and provide a strong approximation for $p_n(x) = n^{d/2}\,p^{*n}(x\sqrt{n})$. In particular, from (2.3)-(2.5) it follows that

$$\int |p_{n1}(x) - p_n(x)|\,dx < 2^{-n},$$

for all n large enough. One of the immediate consequences of this estimate is the bound

$$|v_{n1}(t) - v_n(t)| < 2^{-n} \qquad (t \in \mathbf{R}^d) \tag{2.6}$$

for the characteristic functions $v_n(t) = \int e^{i\langle t,x\rangle} p_n(x)\,dx$ and $v_{n1}(t) = \int e^{i\langle t,x\rangle} p_{n1}(x)\,dx$, corresponding to the densities p_n and p_{n1}.

This property may be sharpened in case of finite moments.

Lemma 2.1. *If* $\mathbf{E}\,|X_1|^s < +\infty$ $(s \geq 0)$*, then for all* n *large enough,*

$$\int (1+|x|^s)\,|\widetilde{p}_n(x) - p_n(x)|\,dx < 2^{-n}.$$

In particular, (2.6) *also holds for all partial derivatives of* v_{n1} *and* v_n *up to order* $m = [s]$.

Proof. By the definition (2.5), $|p_{n1}(x) - p_n(x)| \leq \varepsilon_n(p_{n1}(x) + p_{n2}(x))$, so

$$\int |x|^s \, |p_{n1}(x) - p_n(x)| \, dx \ \leq \ \frac{\varepsilon_n}{1-\varepsilon_n} \, n^{-s/2} \int |x|^s \, \rho_{n1}(x) \, dx + n^{-s/2} \int |x|^s \, \rho_{n2}(x) \, dx.$$

Let $U_1, U_2, \dots$ be independent copies of U and $V_1, V_2, \dots$ be independent copies of V (that are also independent of U_n's), where U and V are random vectors with densities ρ_1 and ρ_2, respectively. From (2.2)

$$\beta_s \equiv \mathbf{E}\,|X_1|^s = (1-b)\,\mathbf{E}\,|U|^s + b\,\mathbf{E}\,|V|^s,$$

so $\mathbf{E}\,|U|^s \leq \beta_s/b$ and $\mathbf{E}\,|V|^s \leq \beta_s/b$ (using $b < \frac{1}{2}$). Therefore, for the normalized sums

$$R_{k,n} = \frac{1}{\sqrt{n}}\,(U_1 + \dots + U_k + V_1 + \dots + V_{n-k}), \qquad 0 \leq k \leq n,$$

we have $\mathbf{E}\,|R_{k,n}|^s \leq \frac{\beta_s}{b}\, n^{s/2}$, if $s \geq 1$, and $\mathbf{E}\,|R_{k,n}|^s \leq \frac{\beta_s}{b}\, n^{1-(s/2)}$, if $0 \leq s \leq 1$. Hence, by the definition of ρ_{n1} and ρ_{n2},

$$\int |x|^s \, \rho_{n1}(x) \, dx \ = \ n^{s/2} \sum_{k=m_0+1}^{n} C_n^k \, (1-b)^k \, b^{n-k} \, \mathbf{E}\,|R_{k,n}|^s \leq \frac{\beta_s}{b} \, n^{s+1},$$

$$\int |x|^s \, \rho_{n2}(x) \, dx \ = \ n^{s/2} \sum_{k=0}^{m_0} C_n^k \, (1-b)^k \, b^{n-k} \, \mathbf{E}\,|R_{k,n}|^s \leq \frac{\beta_s}{b} \, n^{s+1} \, \varepsilon_n.$$

It remains to apply the estimate (2.3) on ε_n, and Lemma 2.1 follows.

We need to extend the assertion of Lemma 2.1 to the relative entropies, which we consider with respect to the standard normal distribution on $\mathbf{R}^d$ with density $\varphi(x) = (2\pi)^{-d/2}\, e^{-|x|^2/2}$. Thus put

$$D_n = \int p_n(x) \log \frac{p_n(x)}{\varphi(x)} \ dx, \qquad \widetilde{D}_n = \int \widetilde{p}_n(x) \log \frac{\widetilde{p}_n(x)}{\varphi(x)} \ dx.$$

Lemma 2.2. *If X_1 has a finite second moment and $D(X_1) < +\infty$, then $|\widetilde{D}_n - D_n| < 2^{-n}$, for all n large enough.*

First, we collect a few elementary properties of the convex function $L(u) = u \log u$ $(u \geq 0)$.

Lemma 2.3. *For all $u, v \geq 0$ and $0 \leq \varepsilon \leq 1$,*

a) $L((1-\varepsilon)\,u + \varepsilon v) \leq (1-\varepsilon)\,L(u) + \varepsilon L(v)\,;$
b) $L((1-\varepsilon)\,u + \varepsilon v) \geq (1-\varepsilon)\,L(u) + \varepsilon L(v) + uL(1-\varepsilon) + vL(\varepsilon)\,;$
c) $L((1-\varepsilon)\,u + \varepsilon v) \geq (1-\varepsilon)\,L(u) - \frac{1}{e}\,u - \frac{1}{e}.$

The first assertion is just Jensen's inequality applied to L. For the second one, for fixed $v \geq 0$ and $0 < \varepsilon < 1$, consider the concave function

$$f(u) = \big[(1-\varepsilon)\,L(u) + \varepsilon L(v) + uL(1-\varepsilon) + vL(\varepsilon)\big] - L((1-\varepsilon)\,u + \varepsilon v), \qquad u \geq 0.$$

Clearly, $f(0) = 0$ and $f'(u) = (1-\varepsilon) \log \frac{(1-\varepsilon)\,u}{(1-\varepsilon)\,u + \varepsilon v} \leq 0$. Hence, $f(u) \leq 0$, thus proving *b)*.

For the last inequality, we use $L(x) \geq -\frac{1}{e}$, for all $x \geq 0$. Let $x = (1-\varepsilon)\,u + \varepsilon v \leq 1$. Then $L((1-\varepsilon)\,u) \leq 0$, so $L(x) \geq L((1-\varepsilon)\,u) - \frac{1}{e}$. This inequality is also true for $x > 1$, since L is increasing and positive in this region. Finally, $L((1-\varepsilon)\,u) = (1-\varepsilon)\,L(u) + uL(1-\varepsilon)$, and $L(1-\varepsilon) \geq -\frac{1}{e}$.

Proof of Lemma 2.2. Assuming that p is (essentially) unbounded, define

$$D_{nj} = \int p_{nj}(x) \log \frac{p_{nj}(x)}{\varphi(x)}\,dx \quad (j = 1,2),$$

so that $\widetilde{D}_n = D_{n,1}$. By Lemma 2.3 *a*), $D_n \leq (1-\varepsilon_n) D_{n1} + \varepsilon_n D_{n2}$. On the other hand, by *b*),

$$D_n \geq \big((1-\varepsilon_n) D_{n1} + \varepsilon_n D_{n2}\big) + \varepsilon_n \log \varepsilon_n + (1-\varepsilon_n)\log(1-\varepsilon_n).$$

In view of (2.3), the two estimates give

$$|D_{n1} - D_n| < C(n + D_{n1} + D_{n2})\, b_1^n, \tag{2.7}$$

which holds for all $n \geq 1$ with some constant C. In addition, by the inequality in *c*) with $\varepsilon = b$, from (2.2) it follows that

$$D(X_1||Z) = \int L\bigg(\frac{p(x)}{\varphi(x)}\bigg)\,\varphi(x)\,dx \geq (1-b)\int \rho_1(x)\log\frac{\rho_1(x)}{\varphi(x)}\,dx - \frac{2}{e},$$

where Z denotes a standard normal random vector in $\mathbf{R}^d$. By the same reason,

$$D(X_1||Z) \geq b\int \rho_2(x)\log\frac{\rho_2(x)}{\varphi(x)}\,dx - \frac{2}{e}.$$

But, if ρ is a density of a random vector U with finite second moment, the relative entropy

$$D(U||Z_{a,\Sigma}) = \int \rho(x)\log\frac{\rho(x)}{\varphi_{a,\Sigma}(x)}\,dx \qquad \big(Z_{a,\Sigma} \sim N(a,\Sigma)\big)$$

is minimized for the mean $a = \mathbf{E}U$ and covariance matrix $\Sigma = \mathrm{Var}(U)$, and is then equal to $D(U)$. Hence, from the above lower bounds on $D(X_1)$ we obtain that

$$D(X_1||Z) \geq (1-b)D(U) - \frac{2}{e}, \qquad D(X_1||Z) \geq bD(V) - \frac{2}{e}, \tag{2.8}$$

where U and V have densities ρ_1 and ρ_2, as in the proof of Lemma 2.1.

Now, by (2.2),

$$\begin{aligned}
\beta^2 \equiv \mathbf{E}\,|X_1|^2 &= (1-b)\,\mathbf{E}\,|U|^2 + b\,\mathbf{E}\,|V|^2 \\
&= (1-b)\,\big(|a_1|^2 + \mathrm{Tr}(\Sigma_1)\big) + b\,\big(|a_2|^2 + \mathrm{Tr}(\Sigma_2)\big) \qquad (\beta > 0),
\end{aligned}$$

where $a_1 = \mathbf{E}\,U$, $a_2 = \mathbf{E}\,V$, and $\Sigma_1 = \mathrm{Var}(U)$, $\Sigma_2 = \mathrm{Var}(V)$. In particular, $|a_1|^2 \leq \beta^2/b$ and $|a_2|^2 \leq \beta^2/b$, and similarly for the traces of the covariance matrices. Note that both U and V have non-degenerate distributions, so the determinants $\sigma_j^2 = \det(\Sigma_j)$ are positive.

Let $U_1, U_2, \dots$ be independent copies of U and $V_1, V_2, \dots$ be independent copies of V (that are also independent of U_n's). Again, by the convexity of the function $u \log u$,

$$D_{n1} \ \leq \ \frac{1}{1-\varepsilon_n} \sum_{k=m_0+1}^{n} C_n^k \, (1-b)^k \, b^{n-k} \int r_{k,n}(x) \, \log \frac{r_{k,n}(x)}{\varphi(x)} \, dx, \tag{2.9}$$

$$D_{n2} \ \leq \ \frac{1}{\varepsilon_n} \sum_{k=0}^{m_0} C_n^k \, (1-b)^k \, b^{n-k} \int r_{k,n}(x) \, \log \frac{r_{k,n}(x)}{\varphi(x)} \, dx, \tag{2.10}$$

where $r_{k,n}$ are densities of the normalized sums $R_{k,n}$ from the proof of Lemma 2.1.

On the other hand, if R is a random vector in $\mathbf{R}^d$ with density $r(x)$, such that $\det\big(\mathrm{Var}(R)\big) = \sigma^2$ $(\sigma > 0)$,

$$D(R||Z) = \int r(x) \log \frac{r(x)}{\varphi(x)} \ dx = D(R) + \log \frac{1}{\sigma^d} + \frac{\mathbf{E}\,|R|^2 - d}{2}. \tag{2.11}$$

In the case $R = R_{k,n}$, we have $\mathbf{E}R = a_1 \frac{k}{\sqrt{n}} + a_2 \frac{n-k}{\sqrt{n}}$, so $|\mathbf{E}R|^2 \ \leq \ \frac{\beta^2}{b}\, n$. Also, $\mathrm{Var}(R) = \frac{k}{n}\,\mathrm{Var}(U) + \frac{n-k}{n}\,\mathrm{Var}(V)$, implying $\mathrm{Tr}(\mathrm{Var}(R)) \leq \frac{\beta^2}{b}$. The two bounds give $\mathbf{E}\,|R|^2 \leq \frac{\beta^2}{b}\,(n+1)$. In addition, by the Minkowski inequality for the determinants of positive definite matrices, $\det\big(\mathrm{Var}(R)\big) \geq \sigma_0^2 = \min(\sigma_1^2, \sigma_2^2)$. Hence, by (2.11),

$$D(R_{k,n}||Z) \leq D(R_{k,n}) + \log \frac{1}{\sigma_0^d} + \frac{\beta^2\,(n+1)}{2b}. \tag{2.12}$$

As for $D(R_{k,n})$, they can be estimated by virtue of the general entropy power inequality

$$e^{2h(X+Y)/d} \geq e^{2h(X)/d} + e^{2h(Y)/d},$$

which holds true for arbitrary independent random vectors in $\mathbf{R}^d$ with finite second moments and absolutely continuous distributions (cf. [Bl], [C-D-T]). It easily implies another general bound $D(X+Y) \leq \max\{D(X), D(Y)\}$. So taking into account (2.8) and using $b < \frac{1}{2}$, we get

$$D(R_{k,n}) \leq \max\{D(U), D(V)\} \leq \frac{1}{b} \left(D(X_1||Z) + \frac{2}{e} \right).$$

Together with (2.12) we arrive at $D(R_{k,n}||Z) \leq C(1 + D(X_1||Z))\,n$ with some constant C, and from (2.9)-(2.10)

$$D_{n1} \leq C(1 + D(X_1||Z))\,n, \qquad D_{n2} \leq C(1 + D(X_1||Z))\,n.$$

Since $D(X_1||Z)$ is finite, it remains to apply (2.7). Lemma 2.2 is proved.

Remark 2.4. If X_1 has a finite second moment and $D(X_1) < +\infty$, the parameter M from (2.1) can be chosen explicitly in terms of b by involving the entropic distance $D(X_1)$ and $\sigma^2 = \det(\Sigma)$, where Σ is the covariance matrix of X_1.

Indeed, putting $a = \mathbf{E}X_1$, we have a simple upper estimate

$$\begin{aligned} \int p \, \log \left(1 + \frac{p}{\varphi_{a,\Sigma}}\right) dx \ &= \ \int \frac{p}{\varphi_{a,\Sigma}} \log \left(1 + \frac{p}{\varphi_{a,\Sigma}}\right) \varphi_{a,\Sigma} \, dx \\ &\leq \ \int \frac{p}{\varphi_{a,\Sigma}} \log \frac{p}{\varphi_{a,\Sigma}} \, dx \ + \ 1 \ = \ D(X_1) + 1. \end{aligned}$$

On the other hand, the original expression majorizes

$$\int_{\{p(x)>M\}} p(x)\, \log \frac{M}{\varphi_{a,\Sigma}(x)}\, dx \ = \ b \log M + \frac{d}{2}\, \log(2\pi e\, \sigma^2),$$

so

$$M \ \leq \ \frac{1}{(2\pi e\, \sigma^2)^{d/2}}\ e^{(D(X_1)+1)/b}.$$

Remark 2.5. If Z_n have absolutely continuous distributions with finite entropies for $n \geq n_0 > 1$, the above construction should properly be modified.

Namely, one may put $\widetilde{p}_n = p_n$, if p_n are bounded, and otherwise apply the same decomposition (2.2) to p_{n_0} in place of p. As a result, for any $n = An_0 + B$ $(A \geq 1,\ 0 \leq B \leq n_0 - 1)$, the partial sum S_n will have the density

$$r_n(x) = \sum_{k=0}^{A} C_A^k\, (1-b)^k\, b^{A-k} \int \big(\rho_1^{*k} * \rho_2^{*(A-k)}\big)(x-y)\, dF_B(y),$$

where F_B is the distribution of S_B. For $A \geq m_0 + 1$, split the above sum into the two parts with summation over $m_0 + 1 \leq k \leq A$ and $0 \leq k \leq m_0$, respectively, so that $r_n = \rho_{n1} + \rho_{n2}$. Then, like in (2.4) and for the same sequence ε_n described in (2.3), define

$$\widetilde{p}_n(x) = \frac{1}{1-\varepsilon_n}\, n^{d/2}\, \rho_{n1}\big(x\sqrt{n}\big).$$

Clearly, these densities are bounded and strongly approximate $p_n(x)$. In particular, for all n large enough, they satisfy the estimates that are simialar to the estimates in Lemmas 2.1-2.2.

3. Edgeworth-type expansions

Let $(X_n)_{n\geq 1}$ be independent identically distributed random variables with mean $\mathbf{E}X_1 = 0$ and variance $\mathrm{Var}(X_1) = 1$. In this section we collect some auxiliary results about Edgeworth-type expansions both for the distribution functions $F_n(x) = \mathbf{P}\{Z_n \leq x\}$ and the densities $p_n(x)$ of the normalized sums $Z_n = S_n/\sqrt{n}$, where $S_n = X_1 + \cdots + X_n$.

If the absolute moment $\mathbf{E}\,|X_1|^s$ is finite for a given $s \geq 2$ and $m = [s]$, define

$$\varphi_m(x) = \varphi(x) + \sum_{k=1}^{m-2} q_k(x)\, n^{-k/2} \tag{3.1}$$

with the functions q_k described in (1.2). Put also

$$\Phi_m(x) = \int_{-\infty}^{x} \varphi_m(y)\, dy = \Phi(x) + \sum_{k=1}^{m-2} Q_k(x)\, n^{-k/2}. \tag{3.2}$$

Similarly to (1.2), the functions Q_k have an explicit description involving the cumulants $\gamma_3, \dots, \gamma_{k+2}$ of X_1. Namely,

$$Q_k(x) \ = \ -\varphi(x) \sum H_{k+2j-1}(x)\, \frac{1}{r_1! \dots r_k!} \left(\frac{\gamma_3}{3!}\right)^{r_1} \dots \left(\frac{\gamma_{k+2}}{(k+2)!}\right)^{r_k},$$

where the summation is carried out over all non-negative integer solutions $(r_1, \dots, r_k)$ to the equation $r_1 + 2r_2 + \cdots + kr_k = k$ with $j = r_1 + \cdots + r_k$ (cf. e.g. [B-RR] or [Pe2] for details).

Theorem 3.1. *Assume that* $\limsup_{|t|\to+\infty} |\mathbf{E}\, e^{itX_1}| < 1$*. If* $\mathbf{E}\,|X_1|^s < +\infty$ $(s \geq 2)$*, then as* $n \to \infty$*, uniformly over all* x

$$(1+|x|^s)\big(F_n(x) - \Phi_m(x)\big) = o\big(n^{-(s-2)/2}\big). \tag{3.3}$$

The implied error term in (3.3) holds uniformly in classes of distributions with prescribed rates of decay of the functions $t \to \mathbf{E}\,|X_1|^s\, 1_{\{|X_1|>t\}}$ and $T \to \sup_{t\geq T} |\mathbf{E}\, e^{itX_1}|$.

For $2 \leq s < 3$ and $m = 2$, there are no terms in the sum (3.2), and then $\Phi_2(x) = \Phi(x)$ is the distribution function of the standard normal law. In this case, (3.3) becomes

$$(1+|x|^s)\big(F_n(x) - \Phi(x)\big) = o\big(n^{-(s-2)/2}\big). \tag{3.4}$$

In fact, in this case Cramer's condition on the characteristic function is not used. The result was obtained by Osipov and Petrov [O-P] (cf. also [Bi] where (3.4) is established with O).

In the case $s \geq 3$ Theorem 2.1 can be found in [Pe2] (Theorem 2, Ch.VI, p. 168). Note that when $s = m$ is integer, the relation (3.3) without the factor $1 + |x|^m$ represents the classical Edgeworth expansion. It is essentially due to Cramér and is described in many papers and textbooks (cf. [E], [F]). However, the case of fractional values of s is more delicate, especially in the next local limit theorem.

Theorem 3.2. *Let* $\mathbf{E}\,|X_1|^s < +\infty$ $(s \geq 2)$*. Suppose* S_{n_0} *has a bounded density for some* n_0*. Then for all* n *large enough,* S_n *have continuous bounded densities* p_n *satisfying, as* $n \to \infty$*,*

$$(1+|x|^m)\,\big(p_n(x) - \varphi_m(x)\big) = o\big(n^{-(s-2)/2}\big) \tag{3.5}$$

uniformly over all x*. Moreover,*

$$(1+|x|^s)\big(p_n(x) - \varphi_m(x)\big) = o\big(n^{-(s-2)/2}\big) + (1+|x|^{s-m})\,\big(O(n^{-(m-1)/2}) + o(n^{-(s-2)})\big). \tag{3.6}$$

In case $2 \leq s < 3$, the last remainder term on the right-hand side is dominating, and (3.6) becomes

$$(1+|x|^s)\big(p_n(x) - \varphi(x)\big) = o\big(n^{-(s-2)/2}\big) + (1+|x|^{s-2})\, o\big(n^{-(s-2)}\big).$$

If $s = m$ is integer and $m \geq 3$, Theorem 3.2 is well-known; (3.5)-(3.6) then simplify to

$$(1+|x|^m)\big(p_n(x) - \varphi_m(x)\big) = o\big(n^{-(m-2)/2}\big). \tag{3.7}$$

In this formulation the result is due to Petrov [Pe1] (cf. [Pe2], p. 211, or [B-RR], p. 192). Without the term $1+|x|^m$, the relation (3.7) goes back to the results of Cramér and Gnedenko (cf. [G-K]).

In the general (fractional) case, Theorem 3.2 has recently been obtained in [B-C-G] by involving the technique of Liouville fractional integrals and derivatives. The assertion (3.6) gives an improvement over (3.5) on relatively large intervals of the real axis, and this is essential in the case of non-integer s.

An obvious weak point in Theorem 3.2 is that it requires the boundedness of the densities p_n, which is, however, necessary for the conclusions, such as (3.5) or (3.7). Nevertheless, this condition may be removed, if we require that (3.5)-(3.6) hold true for slightly modified densities, rather than for p_n.

Theorem 3.3. *Let* $\mathbf{E}\,|X_1|^s < +\infty$ $(s \geq 2)$. *Suppose that, for all* n *large enough,* S_n *have absolutely continuous distributions with densities* p_n. *Then, for some bounded continuous densities* $\widetilde{p}_n$,

a) *the relations* (3.5)-(3.6) *hold true for* $\widetilde{p}_n$ *instead of* p_n ;

b) $\int_{-\infty}^{+\infty} (1+|x|^s)\,|\widetilde{p}_n(x) - p_n(x)|\,dx < 2^{-n}$, *for all* n *large enough*;

c) $\widetilde{p}_n(x) = p_n(x)$ *almost everywhere, if* p_n *is bounded* (*a.e.*)

Here, the property *c*) is added to include Theorem 3.2 in Theorem 3.3 as a particular case. Moreover, one can use the densities $\widetilde{p}_n$ constructed in the previous section with $m_0 = [s]+1$. We refer to [B-C-G] for detailed proofs.

This more general assertion allows us to immediately recover, for example, the central limit theorem with respect to the total variation distance (without the assumption of boundedness of p_n). Namely, we have

$$\|F_n - \Phi_m\|_{\mathrm{TV}} = \int_{-\infty}^{+\infty} |p_n(x) - \varphi_m(x)|\,dx = o\big(n^{-(s-2)/2}\big). \tag{3.8}$$

For $s=2$ and $\varphi_2(x) = \varphi(x)$, this statement corresponds to a theorem of Prokhorov [Pr], while for $s=3$ and $\varphi_3(x) = \varphi(x)\big(1 + \gamma_3\,\frac{x^3-3x}{6\sqrt{n}}\big)$ – to the result of Sirazhdinov and Mamatov [S-M].

Multidimensional case

Similar results are also available in the multidimensional case for integer values $s = m$. In the remaining part of this section, let $(X_n)_{n\geq 1}$ denote independent identically distributed random vectors in the Euclidean space $\mathbf{R}^d$ with mean zero and identity covariance matrix.

Assuming $\mathbf{E}\,|X_1|^m < +\infty$ for some integer $m \geq 2$ (where now $|\cdot|$ denotes the Euclidean norm), introduce the cumulants γ_ν of X_1 and the associated cumulant polynomials $\gamma_k(it)$ up to order m by using the equality

$$\frac{1}{k!}\,\frac{d^k}{du^k}\,\log \mathbf{E}\,e^{iu\langle t,X_1\rangle}\big|_{u=0} = \frac{1}{k!}\,\gamma_k(it) = \sum_{|\nu|=k} \gamma_\nu\,\frac{(it)^\nu}{\nu!} \qquad \big(k=1,\dots,m, \quad t \in \mathbf{R}^d\big).$$

Here the summation runs over all d-tuples $\nu = (\nu_1,\dots,\nu_d)$ with integer components $\nu_j \geq 0$ such that $|\nu| = \nu_1 + \dots + \nu_d = k$. We also write $\nu! = \nu_1!\dots\nu_d!$ and use a standard notation for the generalized powers $z^\nu = z_1^{\nu_1}\dots z_d^{\nu_d}$ of real or complex vectors $z = (z_1,\dots,z_d)$, which are treated as polynomials in z of degree $|\nu|$.

For $1 \leq k \leq m-2$, define the polynomials

$$P_k(it) \;=\; \sum_{r_1+2r_2+\dots+kr_k=k} \frac{1}{r_1!\dots r_k!}\left(\frac{\gamma_3(it)}{3!}\right)^{r_1}\dots\left(\frac{\gamma_{k+2}(it)}{(k+2)!}\right)^{r_k}, \tag{3.9}$$

where the summation is performed over all non-negative integer solutions $(r_1,\dots,r_k)$ to the equation $r_1 + 2r_2 + \dots + kr_k = k$.

Furthermore, like in dimension one, define the approximating functions $\varphi_m(x)$ on $\mathbf{R}^d$ by virtue of the same equality (3.1), where every q_k is determined by its Fourier transform

$$\int e^{i\langle t,x\rangle} q_k(x)\,dx = P_k(it)\,e^{-|t|^2/2}. \tag{3.10}$$

If S_{n_0} has a bounded density for some n_0, then for all n large enough, S_n have continuous bounded densities p_n satisfying (3.7); see [B-RR], Theorem 19.2. We need an extension of this theorem to the case of unbounded densities, as well as integral variants such as (3.8). The first assertion (3.11) in the next theorem is similar to the one-dimensional Theorem 3.3 in the case where $s = m$ is integer, cf. (3.5). For the proof (which we omit), one may apply Lemma 2.1 and follow the standard arguments from [B-RR], Chapter 4.

Theorem 3.4. *Suppose that* $\mathbf{E}\,|X_1|^m < +\infty$ *with some integer* $m \geq 2$. *If, for all* n *large enough,* S_n *have densities* p_n, *then the densities* $\widetilde{p}_n$ *introduced in section* 2 *with* $m_0 = m + 1$ *satisfy*

$$(1 + |x|^m)\,\big(\widetilde{p}_n(x) - \varphi_m(x)\big) = o\big(n^{-(m-2)/2}\big) \tag{3.11}$$

uniformly over all x. *In addition,*

$$\int (1 + |x|^m)\,|\widetilde{p}_n(x) - \varphi_m(x)|\,dx = o\big(n^{-(m-2)/2}\big). \tag{3.12}$$

The second assertion is Theorem 19.5 in [B-RR], where it is stated for $m \geq 3$ under a slightly weaker hypothesis that X_1 has a non-zero absolutely continuous component. Note that, by Lemma 2.1, it does not matter whether $\widetilde{p}_n$ or p_n are present in (3.12).

4. Entropic distance to normality and moderate deviations

Let $X_1, X_2, \dots$ be independent identically distributed random vectors in $\mathbf{R}^d$ with mean zero, identity covariance matrix, and such that $D(Z_n) < +\infty$, for all n large enough.

According to Lemma 2.2 and Remark 2.5, up to an error at most 2^{-n} with large n, the entropic distance to normality, $D_n = D(Z_n)$, is equal to the relative entropy

$$\widetilde{D}_n = \int \widetilde{p}_n(x) \log \frac{\widetilde{p}_n(x)}{\varphi(x)}\,dx,$$

where φ is the density of a standard normal random vector Z in $\mathbf{R}^d$.

Given $T \geq 1$, split the integral into the two parts by writing

$$\widetilde{D}_n = \int_{|x| \leq T} \widetilde{p}_n(x) \log \frac{\widetilde{p}_n(x)}{\varphi(x)}\,dx + \int_{|x| > T} \widetilde{p}_n(x) \log \frac{\widetilde{p}_n(x)}{\varphi(x)}\,dx. \tag{4.1}$$

By Theorems 3.3-3.4, $\widetilde{p}_n$ are uniformly bounded, i.e., $\widetilde{p}_n(x) \leq M$, for all $x \in \mathbf{R}^d$ and $n \geq 1$ with some constant M. Hence, the second integral in (4.1) may be treated from the point of view of moderate deviations (when T is not too large). Indeed, on one hand,

$$\int_{|x|>T} \widetilde{p}_n(x) \log \frac{\widetilde{p}_n(x)}{\varphi(x)}\,dx \leq \int_{|x|>T} \widetilde{p}_n(x) \log \frac{M}{\varphi(x)}\,dx \leq C \int_{|x|>T} |x|^2\,\widetilde{p}_n(x)\,dx,$$

where $C = \frac{1}{2} + \log M + \frac{d}{2}\,\log(2\pi)$. One the other hand, using $u \log u \geq u - 1$, we have a lower bound

$$\int_{|x|>T} \widetilde{p}_n(x) \log \frac{\widetilde{p}_n(x)}{\varphi(x)}\,dx \geq \int_{|x|>T} \big(\widetilde{p}_n(x) - \varphi(x)\big)\,dx \geq -\mathbf{P}\{|Z| > T\}.$$

The two estimates give

$$\left| \int_{|x|>T} \widetilde{p}_n(x) \log \frac{\widetilde{p}_n(x)}{\varphi(x)}\, dx \right| \,\leq\, \mathbf{P}\{|Z|>T\} + C \int_{|x|>T} |x|^2\, \widetilde{p}_n(x)\, dx. \tag{4.2}$$

This is a very general upper bound, valid for any probability density $\widetilde{p}_n$ on $\mathbf{R}^d$, bounded by a constant M (with C as above).

Following (4.1), we are faced with two analytic problems. The first one one is to give a sharp estimate of $|\widetilde{p}_n(x) - \varphi(x)|$ on a relatively large Euclidean ball $|x| \leq T$. Clearly, T has to be small enough, so that results like local limit theorems, such as Theorems 3.2-3.4 may be applied. The second problem is to give a sharp upper bound of the last integral in (4.2). To this aim, we need moderate deviations inequalities, so that Theorems 3.1 and 3.4 are applicable. Anyway, in order to use both types of results we are forced to choose T from a very narrow window only. This value turns out to be approximately

$$T_n = \sqrt{(s-2)\log n + s \log\log n + \rho_n} \qquad (s>2), \tag{4.3}$$

where $\rho_n \to +\infty$ is a sufficiently slowly growing sequence (whose growth will be restricted by the decay of the n-dependent constants in o-expressions of Theorems 3.2-3.4). In case $s=2$, one may put $T_n = \sqrt{\rho_n}$, meaning that $T_n \to +\infty$ is a sufficiently slowly growing sequence.

Lemma 4.1. (The case $d=1$ and s real) *If* $\mathbf{E}X_1 = 0$, $\mathbf{E}X_1^2 = 1$, $\mathbf{E}\,|X_1|^s < +\infty$ $(s \geq 2)$, *then*

$$\int_{|x|>T_n} x^2 \widetilde{p}_n(x)\, dx = o\big((n \log n)^{-(s-2)/2}\big). \tag{4.4}$$

Lemma 4.2. (The case $d \geq 2$ and s integer) *If* X_1 *has mean zero and identity covariance matrix, and* $\mathbf{E}\,|X_1|^m < +\infty$, *then*

$$\int_{|x|>T_n} x^2 \widetilde{p}_n(x)\, dx \,=\, o\big(n^{-(m-2)/2}\, (\log n)^{(m-d)/2}\big) \qquad (m \geq 3), \tag{4.5}$$

and $\int_{|x|>T_n} x^2 \widetilde{p}_n(x)\, dx = o(1)$ *in case* $m=2$.

Note that plenty of results and techniques concerning moderate deviations have been developed by now. Useful estimates can be found e.g. in [G-H]. Restricting ourselves to integer values of $s=m$, one may argue as follows.

Proof of Lemma 4.2. Given $T \geq 1$, write

$$\begin{aligned} \int_{|x|>T} |x|^2\, \widetilde{p}_n(x)\, dx \ &\leq\ \frac{1}{T^{m-2}} \int |x|^m\, \widetilde{p}_n(x)\, dx \\ &\leq\ \frac{1}{T^{m-2}} \int |x|^m\, |\widetilde{p}_n(x) - \varphi_m(x)|\, dx + \frac{1}{T^{m-2}} \int_{|x|>T} |x|^m\, \varphi_m(x)\, dx. \end{aligned} \tag{4.6}$$

By Theorem 3.4, cf. (3.12), the first integral in (4.6) is bounded by $o(n^{-(m-2)/2})$.

From the definition of q_k it follows that $q_k(x) = N(x)\varphi(x)$ with some polynomial N of degree at most $3(m-2)$ (cf. section 6 for details). Hence, from (3.1), $\varphi_m(x) \leq 2\varphi(x)$ on the

balls of large radii $|x| < n^\delta$ with sufficiently large n (where $0 < \delta < \frac{1}{2}$). On the other hand, with some constants C_d, C_d' depending on the dimension, only,

$$\int_{|x|>T} |x|^m \varphi(x)\,dx = C_d \int_T^{+\infty} r^{m+d-1} e^{-r^2/2}\,dr \ \le\ C_d' T^{m+d-2} e^{-T^2/2}. \tag{4.7}$$

But for $T = T_n$ and $s > 2$, we have $e^{-T^2/2} = \frac{1}{T^s}\, o(n^{-(m-2)/2})$, so by (4.6)-(4.7),

$$\int_{|x|>T_n} |x|^2 \widetilde p_n(x)\,dx \ \le\ C\Big(\frac{1}{T^{m-2}} + \frac{1}{T^{m-d}}\Big) o(n^{-(m-2)/2}).$$

Since T_n is of order $\sqrt{\log n}$, (4.5) follows. Also, in the case $m = 2$ (4.6) gives the desired relation

$$\int_{|x|>T_n} |x|^2\, \widetilde p_n(x)\,dx \ \le\ o(1) + \int_{|x|>T_n} |x|^2\, \varphi(x)\,dx \to 0 \qquad (n \to \infty).$$

Proof of Lemma 4.1. The above argument also works for $d = 1$, but it can be refined applying Theorem 3.1 for real s. The case $s = 2$ is already covered, so let $s > 2$.

In view of the decomposition (2.5), integrating by parts, we have, for any $T \ge 0$,

$$(1-\varepsilon_n) \int_{|x|>T} x^2 \widetilde p_n(x)\,dx \ \ \le\ \ \int_{|x|>T} x^2 p_n(x)\,dx \ =\ \int_{|x|>T} x^2\,dF_n(x) \tag{4.8}$$

$$=\ T^2\big(1 - F_n(T) + F_n(-T)\big) + 2\int_T^{+\infty} x\big(1 - F_n(x) + F_n(-x)\big)\,dx, \tag{4.9}$$

where F_n denotes the distribution function of Z_n. (Note that the first inequality in (4.8) should be just ignored in the first case, when p is bounded.)

By (3.3),

$$F_n(x) = \Phi_m(x) + \frac{r_n(x)}{n^{(s-2)/2}}\,\frac{1}{1+|x|^s}, \qquad r_n \ =\ \sup_x |r_n(x)| \to 0 \ \ (n \to \infty).$$

Hence, the first term in (4.9) can be replaced with

$$T^2\big(1 - \Phi_m(T) + \Phi_m(-T)\big) \tag{4.10}$$

at the expense of an error not exceeding (for the values $T \sim \sqrt{\log n}$)

$$\frac{2r_n}{n^{(s-2)/2}}\,\frac{T^2}{1+T^s} \ =\ o\big((n\log n)^{-(s-2)/2}\big). \tag{4.11}$$

Similarly, the integral in (4.9) can be replaced with

$$\int_T^{+\infty} x\big((1 - \Phi_m(x) + \Phi_m(-x)\big)\,dx \tag{4.12}$$

at the expense of an error not exceeding

$$\frac{2r_n}{n^{(s-2)/2}} \int_T^{+\infty} \frac{x\,dx}{1+x^s} \ =\ o\big((n\log n)^{-(s-2)/2}\big). \tag{4.13}$$

To explore the behavior of the expressions (4.10) and (4.12) for $T = T_n$ using precise asymptotics as in (4.3), recall that, by (3.2),

$$1 - \Phi_m(x) = 1 - \Phi(x) - \sum_{k=1}^{m-2} Q_k(x)\, n^{-k/2}.$$

Moreover, we note that $Q_k(x) = N_{3k-1}(x)\,\varphi(x)$, where N_{3k-1} is a polynomial of degree at most $3k-1$. Thus, these functions admit a bound $|Q_k(x)| \leq C_m(1+|x|^{3m})\,\varphi(x)$ with some constants C_m (depending on m and the cumulants $\gamma_3, \dots, \gamma_m$ of X_1), which implies with some other constants

$$|1 - \Phi_m(x)| \leq (1 - \Phi(x)) + \frac{C_m(1+|x|^{3m})}{\sqrt{n}}\,\varphi(x). \tag{4.14}$$

Hence, using $1 - \Phi(x) < \frac{\varphi(x)}{x}$ $(x > 0)$, we get

$$T_n^2\,|1 - \Phi_m(T_n)| \leq CT_n^2\,(1 - \Phi(T_n)) \leq CT_n\,e^{-T_n^2/2} = o\big((n\log n)^{-(s-2)/2}\big). \tag{4.15}$$

A similar bound also holds for $T_n^2\,|\Phi_m(-T_n)|$.

Now, we use (4.14) to estimate (4.12) with $T = T_n$ up to a constant by

$$\int_T^\infty x\,(1 - \Phi(x))\,dx < 1 - \Phi(T) = o\big((n\log n)^{-(s-2)/2}\big).$$

It remains to combine the last relation with (4.11), (4.13) and (4.15). Since $\varepsilon_n \to 0$ in (4.4), Lemma 4.1 follows.

Remark 4.3. Note that the probabilities $\mathbf{P}\{|Z| > T\}$ appearing in (4.2) make a smaller contribution for $T = T_n$ in comparison with the right-hand sides of (4.4)-(4.5). Indeed, we have $\mathbf{P}\{|Z| > T\} \leq C_d\,T^{d-2}e^{-T^2/2}$ $(T \geq 1)$. Hence, the relations (4.4)-(4.5) may be extended to the integrals

$$\int_{|x|>T_n} \widetilde{p}_n(x)\log\frac{\widetilde{p}_n(x)}{\varphi(x)}\,dx.$$

5. **Taylor-type expansion for the entropic distance**

In this section we provide the last auxiliary step towards the proof of Theorem 1.1. In order to describe the multidimensional case, let $X_1, X_2, \dots$ be independent identically distributed random vectors in $\mathbf{R}^d$ with mean zero, identity covariance matrix, and such that $D(Z_{n_0}) < +\infty$ for some n_0.

If p_{n_0} is bounded, then the densities p_n of Z_n $(n \geq n_0)$ are uniformly bounded, and we put $\widetilde{p}_n = p_n$. Otherwise, we use the modified densities $\widetilde{p}_n$ according to the construction of Section 2. In particular, if $\widetilde{Z}_n$ has density $\widetilde{p}_n$, then $|D(\widetilde{Z}_n||Z) - D(Z_n)| < 2^{-n}$ for all n large enough (where Z is a standard normal random vector, cf. Lemma 2.2 and Remark 2.5). Moreover, by Lemmas 4.1-4.2 and Remark 4.3,

$$\left|D(Z_n) - \int_{|x|\leq T_n} \widetilde{p}_n(x)\log\frac{\widetilde{p}_n(x)}{\varphi(x)}\,dx\right| = o(\Delta_n), \tag{5.1}$$

where T_n are defined in (4.3) and

$$\Delta_n = n^{-(s-2)/2} \, (\log n)^{-(s-\max(d,2))/2} \tag{5.2}$$

(with the convention that $\Delta_n = 1$ for the critical case $s = 2$).

Thus, all information about the asymptotic of $D(Z_n)$ is contained in the integral in (5.1). More precisely, writing a Taylor expansion for $\widetilde{p}_n$ using the approximating functions φ_m in Theorems 3.2-3.4 leads to the following representation (which is more convenient in applications such as Corollary 1.2).

Theorem 5.1. *Let $\mathbf{E}\,|X_1|^s < +\infty$ $(s \geq 2)$, assuming that s is integer in case $d \geq 2$. Then*

$$D(Z_n) = \sum_{k=2}^{[\frac{s-2}{2}]} \frac{(-1)^k}{k(k-1)} \int \big(\varphi_m(x) - \varphi(x)\big)^{\,k} \, \frac{dx}{\varphi(x)^{k-1}} + o(\Delta_n) \qquad (m = [s]). \tag{5.3}$$

Note that in case $2 \leq s < 4$ there are no terms in the sum of (5.3) which then simplifies to $D(Z_n) = o(\Delta_n)$.

Proof. In terms of $L(u) = u \log u$ rewrite the integral in (5.1) as

$$\widetilde{D}_{n,1} = \int_{|x| \leq T_n} L\bigg(\frac{\widetilde{p}_n(x)}{\varphi(x)}\bigg) \, \varphi(x)\, dx = \int_{|x| \leq T_n} L\big(1 + u_m(x) + v_n(x)\big) \, \varphi(x)\, dx, \tag{5.4}$$

where

$$u_m(x) = \frac{\varphi_m(x) - \varphi(x)}{\varphi(x)}, \qquad v_n(x) = \frac{\widetilde{p}_n(x) - \varphi_m(x)}{\varphi(x)}.$$

By Theorems 3.3-3.4, more precisely, by (3.6) for $d = 1$, and by (3.11) for $d \geq 2$ and $s = m$ integer, in the region $|x| = O(n^\delta)$ with an appropriate $\delta > 0$, we have

$$|\widetilde{p}_n(x) - \varphi_m(x)| \leq \frac{r_n}{n^{(s-2)/2}} \, \frac{1}{1 + |x|^s}, \qquad r_n \to 0. \tag{5.5}$$

Since $\varphi(x)\,(1 + |x|^s)$ is decreasing for large $|x|$, we obtain that, for all $|x| \leq T_n$,

$$|v_n(x)| \leq C \, \frac{r_n}{n^{(s-2)/2}} \, \frac{e^{T_n^2/2}}{T_n^s} \leq C' \, r_n e^{\rho_n/2}.$$

The last expression tends to zero by a suitable choice of $\rho_n \to \infty$. This will further be assumed. In particular, for n large enough, $|v_n(x)| < \frac{1}{4}$ in $|x| \leq T_n$.

From the definitions of q_k and φ_m, cf. (1.2), (3.1), and (3.10), it follows that

$$|u_m(x)| \leq C_m \, \frac{1 + |x|^{3(m-2)}}{\sqrt{n}} \tag{5.6}$$

with some constants depending on m and the cumulants, only. So, we also have $|u_m(x)| < \frac{1}{4}$ for $|x| \leq T_n$ with sufficiently large n.

Now, by Taylor's formula, for $|u| \leq \frac{1}{4}$, $|v| \leq \frac{1}{4}$,

$$L(1 + u + v) = L(1 + u) + v + \theta_1 \, uv + \theta_2 v^2$$

with some $|\theta_j| \leq 1$ depending on (u,v). Applying this representation with $u = u_m(x)$ and $v = v_n(x)$, we see that $v_n(x)$ can be removed from the right-hand side of (5.4) at the expense of an error not exceeding $|J_1| + J_2 + J_3$, where

$$J_1 = \int_{|x| \leq T_n} \big(\widetilde{p}_n(x) - \varphi_m(x)\big)\,dx, \qquad J_2 = \int_{|x| \leq T_n} |u_m(x)|\,|\widetilde{p}_n(x) - \varphi_m(x)|\,dx,$$

and

$$J_3 = \int_{|x| \leq T_n} \frac{(\widetilde{p}_n(x) - \varphi_m(x))^2}{\varphi(x)}\,dx.$$

But

$$|J_1| = \left|\int_{|x|>T_n} (\widetilde{p}_n(x) - \varphi_m(x))\,dx\right| \leq \int_{|x|>T_n} \widetilde{p}_n(x)\,dx + \int_{|x|>T_n} \varphi_m(x)\,dx. \tag{5.7}$$

By Lemmas 4.1-4.2, the first integral on the right-hand side is T_n^2-times smaller than $o(\Delta_n)$. Also, since $\varphi_m(x) \leq 2\varphi(x)$ for $|x| \leq T_n$ with sufficiently large n, the last integral in (5.7) is bounded by $2\,\mathbf{P}\{|Z| > T_n\} = o(\Delta_n)$, as well. As a result, $J_1 = o(\Delta_n)$.

Applying (5.6) and then the relation (3.12), we conclude as well that

$$J_2 \leq C_m\,\frac{1 + |T_n|^{3(m-2)}}{\sqrt{n}} \int_{|x| \leq T_n} |\widetilde{p}_n(x) - \varphi_m(x)|\,dx = o(\Delta_n).$$

Finally, using (5.5) with $s > 2$, we get up to some constants

$$\begin{aligned} J_3 \ &\leq\ C\,\frac{r_n^2}{n^{s-2}} \int_{|x| \leq T_n} \frac{e^{|x|^2/2}}{1 + |x|^{2s}}\,dx\ \leq\ C_d\,\frac{r_n^2}{n^{s-2}} \int_1^{T_n} r^{d-2s-1}\,e^{r^2/2}\,dr \\ &\leq\ C_d'\,\frac{r_n^2}{n^{s-2}}\,\frac{1}{T_n^{2s-d+2}}\,e^{T_n^2/2}\ =\ o\bigg(\frac{1}{n^{(s-2)/2}\,(\log n)^{(s-d+2)/2}}\bigg)\ =\ o(\Delta_n). \end{aligned}$$

If $s = 2$, all the steps are also valid and give

$$J_3 \leq C_d'\,\frac{r_n^2}{n^{s-2}}\,\frac{1}{T_n^{2s-d+2}}\,e^{T_n^2/2} \to 0$$

for a suitably chosen $T_n \to +\infty$.

Thus, at the expense of an error not exceeding $o(\Delta_n)$ one may remove $v_n(x)$ from (5.4), and then we obtain the relation

$$\widetilde{D}_{n,1} = \int_{|x| \leq T_n} L\big(1 + u_m(x)\big)\,\varphi(x)\,dx + o(\Delta_n), \tag{5.8}$$

which contains specific functions, only.

Moreover, $u_m(x) = u_2(x) = 0$ for $2 \leq s < 3$, and then the theorem is proved.

Next, we consider the case $s \geq 3$. By Taylor's expansion around zero, whenever $|u| < \frac{1}{4}$,

$$L(1+u) = u + \sum_{k=2}^{m-2} \frac{(-1)^k}{k(k-1)}\,u^k + \theta\,u^{m-1}, \qquad |\theta| \leq 1,$$

assuming that the sum has no terms in case $m = 3$. Hence, with some $0 \le \theta \le 1$

$$\int_{|x|\le T_n} L\big(1 + u_m(x)\big)\,\varphi(x)\,dx \ = \ \int_{|x|\le T_n} \big(\varphi_m(x) - \varphi(x)\big)\,dx \tag{5.9}$$

$$+ \sum_{k=2}^{m-2} \frac{(-1)^k}{k(k-1)} \int_{|x|\le T_n} u_m(x)^k\,\varphi(x)\,dx + \theta \int_{\mathbf{R}^d} |u_m(x)|^{m-1}\,\varphi(x)\,dx. \tag{5.10}$$

For n large enough, the second integral in (5.9) has an absolute value

$$\bigg|\int_{|x|>T_n} \big(\varphi_m(x) - \varphi(x)\big)\,dx\bigg| \le \int_{|x|>T_n} \varphi(x)\,dx = \mathbf{P}\{|Z| > T_n\} = o(\Delta_n).$$

This proves the theorem in case $3 \le s < 4$ (when $m = 3$).

Now, let $s \ge 4$. The last integral in (5.10) can be estimated by virtue of (5.6) by

$$\frac{C}{n^{(m-1)/2}} \int_{\mathbf{R}^d} \big(1 + |x|^{3(m-1)(m-2)}\big)\,\varphi(x)\,dx = o(\Delta_n).$$

In addition, the first integral in (5.10) can be extended to the whole space at the expense of an error not exceeding (for all n large enough)

$$\begin{aligned}\int_{|x|>T_n} |u_m(x)|^k\,\varphi(x)\,dx \ &\le \ \frac{C}{n^{k/2}} \int_{|x|>T_n} \big(1 + |x|^{3k(m-2)}\big)\,\varphi(x)\,dx \\ &\le \ \frac{C'\,T_n^{3k(m-2)}}{\sqrt{n}}\ e^{-T_n^2/2} \ = \ o(\Delta_n).\end{aligned}$$

Moreover, if $k > (s-2)/2$,

$$\int |u_m(x)|^k\,\varphi(x)\,dx \le \frac{C}{n^{k/2}} \int \big(1 + |x|^{3k(m-2)}\big)\,\varphi(x)\,dx \ = \ o(\Delta_n).$$

Collecting these estimates in (5.9)-(5.10) and applying them in (5.8), we arrive at

$$\widetilde{D}_{n,1} = \sum_{k=2}^{[\frac{s-2}{2}]} \frac{(-1)^k}{k(k-1)} \int u_m(x)^k\,dx + o(\Delta_n).$$

It remains to apply (5.1). Thus, Theorem 5.1 is proved.

6. Theorem 1.1 and its multidimensional extension

The desired representation (1.3) of Theorem 1.1 can be deduced from Theorem 5.1. Note that the latter covers the multidimensional case as well, although under somewhat stronger moment assumptions.

Thus, let $(X_n)_{n\ge1}$ be independent identically distributed random vectors in $\mathbf{R}^d$ with finite second moment. If the normalized sum $Z_n = (X_1 + \dots + X_n)/\sqrt{n}$ has density $p_n(x)$, the entropic distance to Gaussianity is defined as in dimension one to be the relative entropy

$$D(Z_n) = \int p_n(x)\log\frac{p(x)}{\varphi_{a,\Sigma}(x)}\,dx$$

with respect to the normal law on $\mathbf{R}^d$ with the same mean $a = \mathbf{E}X_1$ and covariance matrix $\Sigma = \mathrm{Var}(X_1)$. This quantity is affine invariant, and in this sense it does not depend on (a, Σ).

Theorem 6.1. *If $D(Z_{n_0}) < +\infty$ for some n_0, then $D(Z_n) \to 0$, as $n \to \infty$. Moreover, given that $\mathbf{E}\,|X_1|^s < +\infty$ $(s \geq 2)$, and that X_1 has mean zero and identity covariance matrix, we have*

$$D(Z_n) = \frac{c_1}{n} + \frac{c_2}{n^2} + \dots + \frac{c_{[(m-2)/2)]}}{n^{[(m-2)/2)]}} + o\big(\Delta_n\big) \qquad (m = [s]), \tag{6.1}$$

where Δ_n are defined in (5.2)*, and where we assume that s is integer in case $d \geq 2$.*

As in Theorem 1.1, here each coefficient c_j is defined according to (1.4). It may be represented as a certain polynomial in the cumulants γ_ν, $3 \leq |\nu| \leq 2j+1$.

Proof. We shall start from the representation (5.3) of Theorem 5.1, so let us return to the definition (3.1),

$$\varphi_m(x) - \varphi(x) = \sum_{r=1}^{m-2} q_r(x)\, n^{-r/2}.$$

In the case $2 \leq s < 3$ (that is, for $m = 2$), the right-hand side contains no terms and is therefore vanishing. Anyhow, raising this sum to the power $k \geq 2$ leads to

$$\big(\varphi_m(x) - \varphi(x)\big)^k = \sum_j n^{-j/2} \sum q_{r_1}(x) \dots q_{r_k}(x),$$

where the inner sum is carried out over all positive integers $r_1, \dots, r_k \leq m-2$ such that $r_1 + \dots + r_k = j$. Respectively, the k-th integral in (5.3) is equal to

$$\sum_j n^{-j/2} \sum \int q_{r_1}(x) \dots q_{r_k}(x)\, \frac{dx}{\varphi(x)^{k-1}}. \tag{6.2}$$

Here the integrals are vanishing for odd j. In dimension one, this follows directly from the definition (1.2) of q_r and the property of the Chebyshev-Hermite polynomials ([Sz])

$$\int_{-\infty}^{+\infty} H_{r_1}(x) \dots H_{r_k}(x)\, \varphi(x)\, dx = 0 \qquad (r_1 + \dots + r_k \text{ is odd}). \tag{6.3}$$

As for the general case, let us look at the structure of the functions q_r. Given a multi-index $\nu = (\nu_1, \dots, \nu_d)$ with integers $\nu_1, \dots, \nu_d \geq 1$, define $H_\nu(x_1, \dots, x_d) = H_{\nu_1}(x_1) \dots H_{\nu_d}(x_d)$, so that

$$\int e^{i\langle t,x\rangle}\, H_\nu(x)\, \varphi(x)\, dx = (it)^\nu\, e^{-|t|^2/2}, \qquad t \in \mathbf{R}^d.$$

Hence, by the definition (3.10),

$$q_r(x) = \varphi(x) \sum_\nu a_\nu H_\nu(x), \tag{6.4}$$

where the coefficients a_ν emerge from the expansion $P_r(it) = \sum_\nu a_\nu (it)^\nu$. Using (3.9), write these polynomials as

$$P_r(it) \ = \sum \frac{1}{l_1! \dots l_r!} \bigg(\sum_{|\nu|=3} \gamma_\nu \, \frac{(it)^\nu}{\nu!} \bigg)^{l_1} \dots \bigg(\sum_{|\nu|=r+2} \gamma_\nu \, \frac{(it)^\nu}{\nu!} \bigg)^{l_r}, \tag{6.5}$$

where the outer summation is performed over all non-negative integer solutions $(l_1, \dots, l_r)$ to the equation $l_1 + 2l_2 + \dots + rl_r = r$. Removing the brackets of the inner sums, we obtain a linear combination of the power polynomials $(it)^\nu$ with exponents of order

$$|\nu| = 3l_1 + \dots + (r+2)l_r = r + 2b_l, \qquad b_l = l_1 + \dots + l_r. \tag{6.6}$$

In particular, $r + 2 \leq |\nu| \leq 3r$, so that $P_r(it)$ is a polynomial of degree at most $3r$, and thus $\varphi_m(x) = N(x)\varphi(x)$, where $N(x)$ is a polynomial of degree at most $3(m-2)$.

Moreover, from (6.4) and (6.6) it follows that

$$\frac{q_{r_1}(x) \dots q_{r_k}(x)}{\varphi(x)^{k-1}} = \varphi(x) \sum a_{\nu^{(1)}} \dots a_{\nu^{(k)}} H_{\nu^{(1)}}(x) \dots H_{\nu^{(k)}}(x), \tag{6.7}$$

where $|\nu^{(1)}| + \dots + |\nu^{(k)}| = r_1 + \dots + r_k \ (\mathrm{mod}\, 2)$. Hence, if $r_1 + \dots + r_k$ is odd, the sum

$$|\nu^{(1)}| + \dots + |\nu^{(k)}| = \sum_{i=1}^d \Big(|\nu_i^{(1)}| + \dots + |\nu_i^{(k)}| \Big)$$

is odd as well. But then at least one of the inner sums, say with coordinate i, must be odd as well. Hence in this case, the integral of (6.7) over x_i will be vanishing by property (6.3).

Thus, in the expression (6.2), only even values of j should be taken into account.

Moreover, since the terms containing $n^{-j/2}$ with $j > s-2$ will be absorbed by the remainder Δ_n in the relation (6.1), we get from (5.3) and (6.2)

$$D(Z_n) \ = \ \sum_{k=2}^{[\frac{s-2}{2}]} \frac{(-1)^k}{k(k-1)} \sum_{\text{even}\, j=2}^{m-2} n^{-j/2} \sum \int q_{r_1}(x) \dots q_{r_k}(x) \, \frac{dx}{\varphi(x)^{k-1}} + o(\Delta_n).$$

Replace now j with $2j$ and rearrange the summation: $D(Z_n) = \sum_{2j \leq m-2} \frac{c_j}{n^j} + o(\Delta_n)$ with

$$c_j = \sum_{k=2}^{m-2} \frac{(-1)^k}{k(k-1)} \sum \int q_{r_1}(x) \dots q_{r_k}(x) \, \frac{dx}{\varphi(x)^{k-1}}.$$

Here the inner summation is carried out over all positive integers $r_1, \dots, r_k \leq m-2$ such that $r_1 + \dots + r_k = 2j$. This implies $k \leq 2j$. Also, $2j \leq m-2$ is equivalent to $j \leq [\frac{s-2}{2}]$. As a result, we arrive at the required relation (6.1) with

$$c_j \ = \ \sum_{k=2}^{2j} \frac{(-1)^k}{k(k-1)} \sum_{r_1 + \dots + r_k = 2j} \int q_{r_1}(x) \dots q_{r_k}(x) \, \frac{dx}{\varphi(x)^{k-1}}. \tag{6.8}$$

Theorem 6.1 and therefore Theorem 1.1 are proved.

Remark. In order to show that c_j is a polynomial in the cumulants γ_ν, $3 \leq |\nu| \leq 2j+1$, first note that $r_1 + \dots + r_k = 2j$, $r_1, \dots, r_k \geq 1$ imply $2j \geq \max_i r_i + (k-1)$, so $\max_i r_i \leq 2j-1$. Thus, the maximal index for the functions q_{r_i} in (6.8) does not exceed $2j-1$. On the other hand,

it follows from (6.4)-(6.5) that P_r and q_r are polynomials in the same set of the cumulants, more precisely, P_r is a polynomials in γ_ν with $3 \leq |\nu| \leq r+2$.

Proof of Corollary 1.2. By Theorem 5.1, cf. (5.3),

$$D(Z_n) = \sum_{k=2}^{m-2} \frac{(-1)^k}{k(k-1)} \int \big(\varphi_m(x) - \varphi(x)\big)^k \, \frac{dx}{\varphi(x)^{k-1}} + o\big(\Delta_n\big). \tag{6.9}$$

Assume that $m \geq 4$ and $\gamma_3 = \cdots = \gamma_{k-1} = 0$ for a given integer $3 \leq k \leq m$. (There is no restriction, when $k = 3$.) Then, by (1.2), $q_1 = \cdots = q_{k-3} = 0$, while $q_{k-2}(x) = \frac{\gamma_k}{k!} H_k(x)\, \varphi(x)$. Hence, according to definition (2.1),

$$\varphi_m(x) - \varphi(x) = \frac{\gamma_k}{k!} \, H_k(x)\varphi(x) \, \frac{1}{n^{(k-2)/2}} + \sum_{j=k-1}^{m-2} \frac{q_j(x)}{n^{j/2}},$$

where the sum is empty in the case $m = 3$. Therefore, the sum in (1.3) will contain powers of $1/n$ starting from $1/n^{k-2}$, and the leading coefficient is due to the quadratic term in (6.9) when $k = 2$. More precisely, if $k - 2 \leq \frac{m-2}{2}$, we get that $c_1 = \cdots = c_{k-3} = 0$, and

$$c_{k-2} = \frac{\gamma_k^2}{2\,k!^{\,2}} \int_{-\infty}^{+\infty} H_k(x)^2 \, \varphi(x)\, dx = \frac{\gamma_k^2}{2\,k!}. \tag{6.10}$$

Hence, if $k \leq \frac{m}{2}$, (6.9) yields $D(Z_n) = \frac{\gamma_k^2}{2\,k!} \, \frac{1}{n^{k-2}} + O\big(n^{-(k-1)}\big)$. Otherwise, the O-term should be replaced by $o((n \log n)^{-(s-2)/2})$. Thus Corollary 1.3 is proved.

By a similar argument, the conclusion may be extended to the multidimensional case. Indeed, if $\gamma_\nu = 0$, for all $3 \leq |\nu| < k$, then by (6.5), $P_1 = \cdots = P_{k-3} = 0$, while

$$P_{k-2}(it) = \sum_{|\nu| = k} \gamma_\nu \, \frac{(it)^\nu}{\nu!}.$$

Correspondingly, in (6.4) we have $q_1 = \cdots = q_{k-3} = 0$ and $q_{k-2}(x) = \varphi(x) \sum_{|\nu|=k} \frac{\gamma_\nu}{\nu!} \, H_\nu(x)$. Therefore,

$$\varphi_m(x) - \varphi(x) \ = \ \varphi(x) \sum_{|\nu|=k} \frac{\gamma_\nu}{\nu!} \, H_\nu(x) \ \frac{1}{n^{(k-2)/2}} \ + \sum_{j=k-1}^{m-2} \frac{q_j(x)}{n^{j/2}}.$$

Applying this relation in (6.9), we arrive at (6.1) with $c_1 = \cdots = c_{k-3} = 0$ and, by orthogonality of the polynomials H_ν,

$$c_{k-2} = \frac{1}{2} \int \bigg(\sum_{|\nu|=k} \frac{\gamma_\nu}{\nu!} \, H_\nu(x) \bigg)^2 \varphi(x)\, dx = \frac{1}{2} \sum_{|\nu|=k} \frac{\gamma_\nu^2}{\nu!}.$$

We may summarize our findings as follows.

Corollary 6.2. *Let $(X_n)_{n \geq 1}$ be i.i.d. random vectors in $\mathbf{R}^d$ $(d \geq 2)$ with mean zero and identity covariance matrix. Suppose that $\mathbf{E}\, |X_1|^m < +\infty$, for some integer $m \geq 4$, and*

$D(Z_{n_0}) < +\infty$, for some n_0. Given $k = 3, 4, \dots, m$, if $\gamma_\nu = 0$ for all $3 \leq |\nu| < k$, we have

$$D(Z_n) \, = \, \frac{1}{2\, n^{k-2}} \sum_{|\nu|=k} \frac{\gamma_\nu^2}{\nu!} + O\Big(\frac{1}{n^{k-1}}\Big) + o\bigg(\frac{1}{n^{(m-2)/2}\, (\log n)^{(m-d)/2}}\bigg). \tag{6.11}$$

The conclusion corresponds to Corollary 1.2, if we replace d with 2 in the remainder on the right-hand side.

As in dimension one, when $\mathbf{E} X_1^{2k} < +\infty$, the o-term may be removed from this representation, while for $k > \frac{m+2}{2}$, the o-term dominates.

When $k = 3$, there is no restriction on the cumulants, and (6.11) becomes

$$D(Z_n) \, = \, \frac{1}{2n} \sum_{|\nu|=3} \frac{\gamma_\nu^2}{\nu!} + O\Big(\frac{1}{n^2}\Big) + o\bigg(\frac{1}{n^{(m-2)/2}\, (\log n)^{(m-d)/2}}\bigg).$$

If $\mathbf{E}\,|X_1|^4 < +\infty$, we get $D(Z_n) = O(1/n)$ for $d \leq 4$, and only $D(Z_n) = o((\log n)^{(d-4)/2}/n)$ for $d \geq 5$. However, if $\mathbf{E}\,|X_1|^5 < +\infty$, we always have $D(Z_n) = O(1/n)$ regardless of the dimension d.

Technically, this slight difference between conclusions for different dimensions is due to the dimension-dependent asymptotic $\int_{|x|>T} |x|^2 \, \varphi(x)\, dx \sim C_d \, T^d e^{-T^2/2}$.

7. **Convolutions of mixtures of normal laws**

Is the asymptotic description of $D(Z_n)$ in Theorem 1.1 still optimal, if no expansion terms of order n^{-j} are present? This is exactly the case for $2 \leq s < 4$.

In order to answer the question, we examine one special class of probability distributions that can be described as mixtures of normal laws on the real line with mean zero. They have densities of the form

$$p(x) = \int_0^{+\infty} \varphi_\sigma(x) \, dP(\sigma) \qquad (x \in \mathbf{R}), \tag{7.1}$$

where P is a (mixing) probability measure on the positive half-axis $(0, +\infty)$, and where

$$\varphi_\sigma(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-x^2/(2\sigma^2)}$$

is the density of the normal law with mean zero and variance σ^2 (As usual, we write $\varphi(x)$ in the standard normal case with $\sigma = 1$).

Equivalently, let $p(x)$ denote the density of the random variable $X_1 = \rho Z$, where the factors $Z \sim N(0,1)$ and $\rho > 0$, having the distribution P, are independent. Such distributions appear naturally, for example, as limit laws in randomized models of summation (cf. e.g. [B-G]).

For densities such as (7.1), we need a refinement of the local limit theorem for convolutions, described in the expansions (3.5)-(3.6). More precisely, our aim is to find a representation with an essentially smaller remainder term compared to $o(n^{-(s-2)/2})$.

Thus, let $X_1, X_2, \dots$ be independent random variables, having a common density $p(x)$ as in (7.1), and let $p_n(x)$ denote the density of the normalized sum $Z_n = (X_1 + \dots + X_n)/\sqrt{n}$. If

$X_1 = \rho Z$, where $Z \sim N(0,1)$ and $\rho > 0$ are independent, then $\mathbf{E}X_1^2 = \mathbf{E}\rho^2$ and more generally

$$\mathbf{E}\,|X_1|^s = \beta_s\,\mathbf{E}\,\rho^s = \beta_s \int_0^{+\infty} \sigma^s\,dP(\sigma),$$

where β_s denotes the s-th absolute moment of Z.

Note that $p(x)$ is unimodal with mode at the origin, and $p(0) = \mathbf{E}\frac{1}{\rho\sqrt{2\pi}}$. If $\rho \geq \sigma_0 > 0$, the density is bounded, and therefore the entropy $h(X_1)$ is finite.

Proposition 7.1. *Assume that* $\mathbf{E}\rho^2 = 1$, $\mathbf{E}\rho^s < +\infty$ $(2 < s \leq 4)$. *If* $\mathbf{P}\{\rho \geq \sigma_0\} = 1$ *with some constant* $\sigma_0 > 0$, *then uniformly over all* x,

$$p_n(x) = \varphi(x) + n \int_0^{+\infty} \big(\varphi_{\sigma_n}(x) - \varphi(x)\big)\,dP(\sigma) + O\Big(\frac{1}{n^{s-2}}\Big), \tag{7.2}$$

where $\sigma_n = \sqrt{1 + \frac{\sigma^2-1}{n}}$.

Of course, when $\mathbf{E}\rho^s < +\infty$ for $s > 4$, the proposition may still be applied, but with $s = 4$. In this case (7.2) has a remainder term of order $O(\frac{1}{n^2})$.

Note that $p_n(x)$ may also be described as the density of $Z_n = \sqrt{\frac{\rho_1^2+\dots+\rho_n^2}{n}}\,Z$, where ρ_k are independent copies of ρ (independent of Z as well). This represention already indicates the closeness of p_n and φ and suggests to appeal to the law of large numbers. However, we shall choose a different approach based on the study of the characteristic functions of Z_n.

Obviously, the characteristic function of X_1 is given by

$$v(t) = \mathbf{E}\,e^{itX_1} = \mathbf{E}\,e^{-\rho^2 t^2/2} \qquad (t \in \mathbf{R}).$$

Using Jensen's inequality and the assumption $\rho \geq \sigma_0 > 0$, we get a two-sided estimate

$$e^{-t^2/2} \leq v(t) \leq e^{-\sigma_0^2 t^2/2}. \tag{7.3}$$

In particular, the function $\psi(t) = e^{t^2/2}v(t) - 1$ is non-negative for all t real.

Lemma 7.2. *If* $\mathbf{E}\rho^2 = 1$, $M_s = \mathbf{E}\rho^s < +\infty$ $(2 \leq s \leq 4)$, *then for all* $|t| \leq 1$,

$$0 \leq \psi(t) \leq M_s|t|^s.$$

Proof. We may assume $0 < t \leq 1$. Write $\psi(t) = \mathbf{E}\,\big(e^{-(\rho^2-1)\,t^2/2} - 1\big)$. The expression under the expectation sign is non-positive for $\rho t > 1$, so

$$\psi(t) \leq \mathbf{E}\,\big(e^{-(\rho^2-1)\,t^2/2} - 1\big)1_{\{\rho \leq 1/t\}}.$$

Let $x = -(\rho^2-1)\,t^2$. Clearly, $|x| \leq 1$ for $\rho \leq 1/t$. Using $e^x \leq 1 + x + x^2$ $(|x| \leq 1)$ and $\mathbf{E}\rho^2 = 1$, we get

$$\begin{aligned}
\psi(t) \ &\leq\ -\frac{t^2}{2}\,\mathbf{E}\,(\rho^2-1)\,1_{\{\rho \leq 1/t\}} + \frac{t^4}{4}\,\mathbf{E}\,(\rho^2-1)^2\,1_{\{\rho \leq 1/t\}} \\
&=\ \frac{t^2}{2}\,\mathbf{E}\,(\rho^2-1)\,1_{\{\rho > 1/t\}} + \frac{t^4}{4}\,\mathbf{E}\,(\rho^2-1)^2\,1_{\{\rho \leq 1/t\}}.
\end{aligned} \tag{7.4}$$

The last expectation is equal to

$$\begin{aligned}\mathbf{E}\,\rho^4\,1_{\{\rho\le 1/t\}} - 2\,\mathbf{E}\,(1-\rho^2)\,1_{\{\rho>1/t\}} + \mathbf{P}\{\rho\le 1/t\} \ &\le\ \mathbf{E}\,\rho^4\,1_{\{\rho\le 1/t\}} + 2\,\mathbf{E}\,\rho^2\,1_{\{\rho>1/t\}} - 1 \\ &\le\ \mathbf{E}\,\rho^4\,1_{\{\rho\le 1/t\}} + \mathbf{E}\,\rho^2\,1_{\{\rho>1/t\}}.\end{aligned}$$

Together with (7.4), this gives

$$\psi(t) \le \frac{3t^2}{4}\,\mathbf{E}\,\rho^2\,1_{\{\rho>1/t\}} + \frac{t^4}{4}\,\mathbf{E}\,\rho^4\,1_{\{\rho\le 1/t\}}. \tag{7.5}$$

Finally, $\mathbf{E}\rho^2\,1_{\{\rho>1/t\}} \le \mathbf{E}\rho^s\,t^{s-2}\,1_{\{\rho>1/t\}} \le M_s t^{s-2}$ and $\mathbf{E}\rho^4\,1_{\{\rho\le 1/t\}} \le \mathbf{E}\rho^s\,t^{s-4}\,1_{\{\rho\le 1/t\}} \le M_s t^{s-4}$. It remains to use these estimates in (7.5) and Lemma 7.2 is proved.

Proof of Proposition 7.1. The characteristic functions $v_n(t) = v(\frac{t}{\sqrt{n}})^n$ of Z_n are real-valued and admit, by (7.3), similar bounds

$$e^{-t^2/2} \le v_n(t) \le e^{-\sigma_0^2 t^2/2}. \tag{7.6}$$

In particular, one may apply the inverse Fourier transform to represent the density of Z_n as

$$p_n(x) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-itx} v_n(t)\,dt = \frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-itx-t^2/2}\big(1+\psi(t/\sqrt{n})\big)^n\,dt.$$

Letting $T_n = \frac{4}{\sigma_0}\,\log n$, we split the integral into the two regions, defined by

$$I_1 = \int_{|t|\le T_n} e^{-itx} v_n(t)\,dt, \qquad I_2 = \int_{|t|>T_n} e^{-itx} v_n(t)\,dt.$$

By the upper bound in (7.6),

$$|I_2| \le \int_{|t|>T_n} e^{-\sigma_0^2 t^2/2}\,dt < \frac{2}{\sigma_0}\,e^{-\sigma_0^2 T_n^2/2} = \frac{2}{\sigma_0\,n^8}. \tag{7.7}$$

In the interval $|t|\le T_n$, by Lemma 7.2, $\psi(\frac{t}{\sqrt{n}}) \le \frac{M_s\,|t|^s}{n^{s/2}} \le \frac{1}{n}$, for all $n \ge n_0$. But for $0 \le \varepsilon \le \frac{1}{n}$, there is a simple estimate $0 \le (1+\varepsilon)^n - 1 - n\varepsilon \le 2\,(n\varepsilon)^2$. Hence, once more by Lemma 7.2,

$$0 \ \le\ \big(1+\psi(t/\sqrt{n})\big)^n - 1 - n\psi(t/\sqrt{n}) \ \le\ 2\,\big(n\psi(t/\sqrt{n})\big)^2 \ \le\ 2M_s^2\,\frac{|t|^{2s}}{n^{s-2}} \qquad (n \ge n_0).$$

This gives

$$\bigg|I_1 - \int_{|t|\le T_n} e^{-itx-t^2/2}\big(1+n\psi(t/\sqrt{n})\big)\,dt\bigg| \le \frac{2M_s^2}{n^{s-2}}\int_{-\infty}^{+\infty} |t|^{2s}\,e^{-t^2/2}\,dt. \tag{7.8}$$

In addition,

$$\bigg|\int_{|t|>T_n} e^{-itx-t^2/2}\big(1+n\psi(t/\sqrt{n})\big)\,dt\bigg| \le \int_{|t|>T_n} e^{-t^2/2}\,dt + n\int_{|t|>T_n} e^{-t^2/2}\psi(t/\sqrt{n})\,dt.$$

Here, the first integral on the right-hand side is of order $O(n^{-8})$. To estimate the second one, recall that, by (7.3), $\psi(t) = e^{t^2/2}v(t) - 1 \le e^{(1-\sigma_0^2)t^2/2}$. Hence, $\psi(t/\sqrt{n}) \le e^{(1-\sigma_0^2)t^2/2}$ and

$$\int_{|t|>T_n} e^{-t^2/2}\psi(t/\sqrt{n})\,dt \le \int_{|t|>T_n} e^{-\sigma_0^2 t^2/2}\,dt < \frac{2}{\sigma_0\,n^8}.$$

Together with (7.7) and (7.8) these bounds imply that

$$p_n(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-itx-t^2/2} \big(1 + n\psi(t/\sqrt{n})\big)\, dt + O\Big(\frac{1}{n^{s-2}}\Big)$$

uniformly over all x. It remains to note that

$$\begin{aligned} \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-itx-t^2/2} \psi(t/\sqrt{n})\, dt &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-itx-t^2/2} \big(e^{t^2/2n} v(t/\sqrt{n}) - 1\big)\, dt \\ &= \int_0^{+\infty} \big(\varphi_{\sigma_n}(x) - \varphi(x)\big)\, dP(\sigma). \end{aligned}$$

Proposition 7.1 is proved.

Remark 7.3. An inspection of (7.5) shows that, in the case $2 < s < 4$, Lemma 7.2 may slightly be sharpened to $\psi(t) = o(|t|^s)$. Correspondingly, the O-relation in Proposition 7.1 can be replaced with an o-relation. This improvement is convenient, but not crucial for the proof of Theorem 1.3.

8. **Lower bounds. Proof of Theorem 1.3**

Let $X_1, X_2, \dots$ be independent random variables with a common density of the form

$$p(x) = \int_0^{+\infty} \varphi_\sigma(x)\, dP(\sigma), \quad x \in \mathbf{R}.$$

Equivalently, let $X_1 = \rho Z$ with independent random variables $Z \sim N(0,1)$ and $\rho > 0$ having distribution P.

A basic tool for proving Theorem 1.3 will be the following lower bound on the entropic distance to Gaussianity for the partial sums $S_n = X_1 + \dots + X_n$.

Proposition 8.1. *Let $\mathbf{E}\rho^2 = 1$, $\mathbf{E}\rho^s < +\infty$ $(2 < s < 4)$, and $\mathbf{P}\{\rho \geq \sigma_0\} = 1$ with $\sigma_0 > 0$. Assume that, for some $\gamma > 0$,*

$$\liminf_{n \to \infty} n^{s-\frac{1}{2}} \int_{n^{\frac{1}{2}+\gamma}}^{+\infty} \frac{1}{\sigma}\, dP(\sigma) > 0. \tag{8.1}$$

Then with some absolute constant $c > 0$ and some constant $\delta > 0$

$$D(S_n) \ \geq \ c\, n \log n \ \mathbf{P}\big\{\rho \geq \sqrt{n \log n}\,\big\} + o\bigg(\frac{1}{n^{\frac{s-2}{2}+\delta}}\bigg). \tag{8.2}$$

In fact, in (8.2) one may take any positive number $\delta < \min\{\gamma s, \frac{s-2}{2}\}$.

Proof. By Proposition 7.1 and Remark 7.3, uniformly over all x,

$$p_n(x) = \varphi(x) + n \int_0^{+\infty} \big(\varphi_{\sigma_n}(x) - \varphi(x)\big)\, dP(\sigma) + o\Big(\frac{1}{n^{s-2}}\Big), \tag{8.3}$$

where p_n is the density of $S_n/\sqrt{n}$ and $\sigma_n = \sqrt{1 + \frac{\sigma^2 - 1}{n}}$.

Define the sequence

$$N_n = \frac{n^{\frac{1}{2}+\gamma}}{5\sqrt{\log n}}$$

for n large enough (so that $N_n \geq 1$). By Chebyshev's inequality,

$$\mathbf{P}\{\rho \geq N_n\} \leq 5^s M_s \frac{\log^2 n}{n^{(\frac{1}{2}+\gamma)s}} = o\bigg(\frac{1}{n^{\frac{s}{2}+\delta}}\bigg), \qquad 0<\delta<\gamma s. \tag{8.4}$$

Using $u\log u \geq u-1$ $(u \geq 0)$ and applying (8.3), we may write

$$\begin{aligned} I_n \equiv \int_{|x|\leq 4\sqrt{\log n}} p_n(x) \log\frac{p_n(x)}{\varphi(x)}\,dx \ &\geq \ \int_{|x|\leq 4\sqrt{\log n}} \big(p_n(x)-\varphi(x)\big)\,dx \\ &\geq \ n\int_0^{+\infty}\int_{|x|\leq 4\sqrt{\log n}} \big(\varphi_{\sigma_n}(x)-\varphi(x)\big)\,dx\,dP(\sigma) - C\,\frac{\sqrt{\log n}}{n^{s-2}} \end{aligned} \tag{8.5}$$

with some constant C.

Note that $\sigma_n < 1$ for $\sigma < 1$, and then, for any $T > 0$,

$$\int_{|x|\leq T} \big(\varphi_{\sigma_n}(x)-\varphi(x)\big)\,dx = 2\,\big(\Phi(T/\sigma_n)-\Phi(T)\big) > 0,$$

where Φ denotes the distribution function of the standard normal law. Hence, the outer integral in (8.5) may be restricted to the values $\sigma \geq 1$. Moreover, by (8.4), one may also restrict this integral to the values $\sigma \geq N_n$. More precisely, (8.4) gives

$$n\,\bigg|\int_{N_n}^{+\infty}\int_{|x|\leq 4\sqrt{\log n}} \big(\varphi_{\sigma_n}(x)-\varphi(x)\big)\,dx\,dP(\sigma)\bigg| \ \leq\ n\mathbf{P}\{\rho\geq N_n\} \ =\ o\bigg(\frac{1}{n^{\frac{s-2}{2}+\delta}}\bigg).$$

Comparing this relation with (8.5) and imposing the additional requirement $\delta < \frac{s-2}{2}$, we get

$$\begin{aligned} I_n \ &\geq\ n\int_1^{N_n}\int_{|x|\leq 4\sqrt{\log n}} \big(\varphi_{\sigma_n}(x)-\varphi(x)\big)\,dx\,dP(\sigma) + o\bigg(\frac{1}{n^{\frac{s-2}{2}+\delta}}\bigg) \\ &=\ -2n\int_1^{N_n}\int_{\frac{4}{\sigma_n}\sqrt{\log n}}^{4\sqrt{\log n}} \varphi(x)\,dx\ dP(\sigma) + o\bigg(\frac{1}{n^{\frac{s-2}{2}+\delta}}\bigg). \end{aligned} \tag{8.6}$$

Now, let us estimate from below $p_n(x)$ in the region $4\sqrt{\log n} \leq |x| \leq n^\gamma$. If $|x| \geq 4\sqrt{\log n}$, it follows from (8.3) that

$$p_n(x) = n\int_0^{+\infty} \varphi_{\sigma_n}(x)\,dP(\sigma) + o\bigg(\frac{1}{n^{s-2}}\bigg). \tag{8.7}$$

Consider the function

$$g_n(x) = \int_0^{+\infty} \frac{\varphi_{\sigma_n}(x)}{\varphi(x)}\,dP(\sigma).$$

Note that $1 \leq \sigma_n \leq \sigma$ for $\sigma \geq 1$. In this case, the ratio $\frac{\varphi_{\sigma_n}(x)}{\varphi(x)}$ is non-increasing in $x \geq 0$. Moreover, for $\sigma \geq \sqrt{3n+1}$, we have $\sigma_n^2 = 1+\frac{\sigma^2-1}{n} \geq 4$, so $1-\frac{1}{\sigma_n^2} \geq \frac{3}{4}$. Hence, for $|x| \geq 4\sqrt{\log n}$,

$$\frac{\varphi_{\sigma_n}(x)}{\varphi(x)} = \frac{1}{\sigma_n}\,e^{\frac{x^2}{2}\,(1-\frac{1}{\sigma_n^2})} \geq \frac{n^6}{\sigma}.$$

Therefore,

$$g_n(x) \geq n^6 \int_{4\sqrt{\log n}}^{+\infty} \frac{1}{\sigma}\, dP(\sigma).$$

But by the assumption (8.1), the last expression tends to infinity with n, so for all n large enough, $g_n(x) \geq 2$ in the interval $|x| \geq 4\sqrt{\log n}$.

Next, if $\sigma \geq |x|\sqrt{n}$, then $\sigma_n^2 = 1 + \frac{\sigma^2-1}{n} \geq x^2$, so $\frac{x^2}{2\sigma_n^2} \leq \frac{1}{2}$. On the other hand,

$$\sigma_n^2 < 1 + \frac{\sigma^2}{n} = \frac{n+\sigma^2}{n} \leq \frac{\frac{\sigma^2}{x^2} + \sigma^2}{n} \leq \frac{2\sigma^2}{n},$$

since $|x| \geq 4\log n > 1$ for $n \geq 2$. The two estimates give

$$\varphi_{\sigma_n}(x) = \frac{1}{\sigma_n\sqrt{2\pi}}\, e^{-x^2/2\sigma_n^2} \geq \frac{\sqrt{n}}{6\sigma}.$$

Therefore, whenever $4\sqrt{\log n} \leq |x| \leq n^\gamma$,

$$n\int_0^{+\infty} \varphi_{\sigma_n}(x)\, dP(\sigma) \ \geq\ \frac{n^{3/2}}{6}\int_{x\sqrt{n}}^{+\infty} \frac{1}{\sigma}\, dP(\sigma) \ \geq\ \frac{n^{3/2}}{6}\int_{n^{\frac{1}{2}+\gamma}}^{+\infty} \frac{1}{\sigma}\, dP(\sigma).$$

By the assumption (8.1), the last expression and therefore the left integral are larger than $\frac{c}{n^{s-2}}$ with some constant $c > 0$. Consequently, the remainder term in (8.7) is indeed smaller, so that for all n large enough, we may write, for example,

$$p_n(x) \ \geq\ 0.8\, n\int_0^{+\infty} \varphi_{\sigma_n}(x)\, dP(\sigma) \ =\ 0.8\, n\ g_n(x)\varphi(x) \qquad \left(4\sqrt{\log n} \leq |x| \leq n^\gamma\right).$$

Since $g_n(x) \geq 2$ for $|x| \geq 4\sqrt{\log n}$ with large n, we have in this region $\frac{p_n(x)}{\varphi(x)} \geq 1.6\, n > n$, so

$$p_n(x)\log\frac{p_n(x)}{\varphi(x)} \ \geq\ p_n(x)\log n \ \geq\ 0.8\, n\log n\int_0^{+\infty} \varphi_{\sigma_n}(x)\, dx\ dP(\sigma).$$

Hence,

$$\begin{aligned}\int_{4\sqrt{\log n}\leq|x|\leq n^\gamma} p_n(x)\log\frac{p_n(x)}{\varphi(x)}\, dx \ &\geq\ 0.8\, n\log n\int_0^{+\infty}\int_{4\sqrt{\log n}\leq|x|\leq n^\gamma} \varphi_{\sigma_n}(x)\, dx\ dP(\sigma) \\ &=\ 1.6\, n\log n\int_0^{+\infty}\int_{\frac{4}{\sigma_n}\sqrt{\log n}}^{\frac{n^\gamma}{\sigma_n}} \varphi(x)\, dx\ dP(\sigma). \qquad (8.8)\end{aligned}$$

At this point, it is useful to note that $\frac{n^\gamma}{\sigma_n} \geq 4\sqrt{\log n}$, as long as $\sigma \leq N_n$ with n large enough. Indeed, in this case $\sigma_n^2 \leq (1-\frac{1}{n}) + \frac{N_n^2}{n} < 1 + \frac{n^{2\gamma}}{25\,\log n}$, so

$$\left(4\sigma_n\sqrt{\log n}\right)^2 \ \leq\ 16\log n\left(1 + \frac{n^{2\gamma}}{25\,\log n}\right) < n^{2\gamma},$$

for all n large enough. Hence, from (8.8),

$$\int_{4\sqrt{\log n}\leq|x|\leq n^\gamma} p_n(x)\log\frac{p_n(x)}{\varphi(x)}\, dx \ \geq\ 1.6\, n\log n\int_0^{N_n}\int_{\frac{4}{\sigma_n}\sqrt{\log n}}^{4\sqrt{\log n}} \varphi(x)\, dx\ dP(\sigma).$$

But the last expression dominates the double integral in (8.6) with a factor of $2n$. Therefore, combining the above estimate with (8.6), we get

$$\int_{|x|\le n^\gamma} p_n(x)\log\frac{p_n(x)}{\varphi(x)}\,dx \ \ge\ 1.4\,n\log n\int_0^{N_n}\int_{\frac{4}{\sigma_n}\sqrt{\log n}}^{4\sqrt{\log n}}\varphi(x)\,dx\ dP(\sigma)+o\Big(\frac{1}{n^{\frac{s-2}{2}+\delta}}\Big).$$

Finally, we may extend the outer integral on the right-hand side to all values $\sigma>0$ by noting that, by (8.4),

$$n\log n\int_{N_n}^{+\infty}\int_{\frac{4}{\sigma_n}\sqrt{\log n}}^{4\sqrt{\log n}}\varphi(x)\,dx\ dP(\sigma)\ \le\ n\log n\ \mathbf{P}\{\rho>N_n\}\ =\ o\Big(\frac{1}{n^{\frac{s-2}{2}+\delta}}\Big).$$

Hence,

$$\int_{|x|\le n^\gamma} p_n(x)\log\frac{p_n(x)}{\varphi(x)}\,dx \ \ge\ 1.4\,n\log n\int_0^{+\infty}\int_{\frac{4}{\sigma_n}\sqrt{\log n}}^{4\sqrt{\log n}}\varphi(x)\,dx\ dP(\sigma)+o\Big(\frac{1}{n^{\frac{s-2}{2}+\delta}}\Big). \tag{8.9}$$

For the remaining values $|x|\ge n^\gamma$, one can just use the property $u\log u\ge -\frac{1}{e}$ to get a simple lower bound

$$\begin{aligned}\int_{|x|>n^\gamma} p_n(x)\log\frac{p_n(x)}{\varphi(x)}\,dx &\ \ge\ \int_{|x|>n^\gamma,\,p_n(x)\le\varphi(x)} p_n(x)\log\frac{p_n(x)}{\varphi(x)}\,dx\\ &\ \ge\ -\frac{1}{e}\int_{|x|>n^\gamma,\,p_n(x)\le\varphi(x)}\varphi(x)\,dx\ \ge\ -e^{-n^{2\gamma}/2}.\end{aligned}$$

Together with (8.9) this yields

$$\int_{-\infty}^{+\infty} p_n(x)\log\frac{p_n(x)}{\varphi(x)}\,dx \ \ge\ 1.4\,n\log n\int_0^{+\infty}\int_{\frac{4}{\sigma_n}\sqrt{\log n}}^{4\sqrt{\log n}}\varphi(x)\,dx\ dP(\sigma)+o\Big(\frac{1}{n^{\frac{s-2}{2}+\delta}}\Big).$$

To simplify, finally note that $\frac{4}{\sigma_n}\sqrt{\log n}\le 4$ for $\sigma\ge\sqrt{n\log n}$. In this case the last integral is separated from zero (for large n), hence with some absolute constant $c>0$

$$\int_{-\infty}^{+\infty} p_n(x)\log\frac{p_n(x)}{\varphi(x)}\,dx\ \ge\ c\,n\log n\ \mathbf{P}\big\{\rho\ge\sqrt{n\log n}\big\}+o\Big(\frac{1}{n^{\frac{s-2}{2}+\delta}}\Big).$$

This is exactly the required inequality (8.2) and Proposition 8.1 is proved.

Proof of Theorem 1.3. Given $\eta>0$, one may apply Proposition 8.1 to the probability measure P with density

$$\frac{dP(\sigma)}{d\sigma}=\frac{c_\eta}{\sigma^{s+1}(\log\sigma)^\eta},\qquad \sigma>2,$$

and extending it to an interval $[\sigma_0,2]$ to meet the requirement $\int_{\sigma_0}^{+\infty}\sigma^2\,dP(\sigma)=1$ (with some $0<\sigma_0<1$ and a normalizing constant $c_\eta>0$). It is easy to see that in this case the condition (8.1) is fulfilled for $0<\gamma\le\frac{s-2}{2s}$. In addition, if ρ has the distribution P, we have

$$\mathbf{P}\{\rho\ge\sigma\}\ge\mathrm{const}\,\frac{1}{\sigma^s(\log\sigma)^\eta},$$

for all σ large enough. Hence, by taking $\sigma=\sqrt{n\log n}$, (8.2) provides the desired lower bound.

Remark. In case $s = 2$ (that is, with minimal moment assumptions), the mixtures of the normal laws with discrete mixing measures P were used by Matskyavichyus [M] in the central limit theorem in terms of the Kolmogorov distance. Namely, it is shown that, for any prescribed sequence $\varepsilon_n \to 0$, one may choose P such that $\Delta_n = \sup_x |F_n(x) - \Phi(x)| \geq \varepsilon_n$ for all n large enough (where F_n is the distribution function of Z_n). In view of the Pinsker-type inequality, one may conclude that

$$D(Z_n) \geq \frac{1}{2}\,\Delta_n^2 \geq \frac{1}{2}\,\varepsilon_n^2.$$

Therefore, $D(Z_n)$ may decay arbitrarily slow.

References

[A-B-B-N1] S. Artstein, K.M. Ball, F. Barthe and A. Naor. Solution of Shannon's problem on the monotonicity of entropy. J. Amer. Math. Soc., 17 (2004), 975-982.

[A-B-B-N2] S. Artstein, K.M. Ball, F. Barthe and A. Naor. On the rate of convergence in the entropic central limit theorem. Probab. Theory Related Fields, 129 (2004), no. 3, 381-390.

[Ba] A.R. Barron. Entropy and the central limit theorem. Ann. Probab. 14 (1986), no. 1, 336-342.

[B-J] A.R. Barron and O. Johnson. Fisher information inequalities and the central limit theorem. Probab. Theory Related Fields, 129 (2004), no. 3, 391-409.

[B-M] A.R. Barron and M. Madiman. Generalized entropy power inequalities and monotonicity properties of information. IEEE Trans. Inform. Theory, 53 (2007), no. 7, 2317-2329.

[B-RR] R.N. Bhattacharya and R. Ranga Rao. Normal approximation and asymptotic expansions. John Wiley & Sons, Inc. 1976. Also: Soc. for Industrial and Appl. Math., Philadelphia, 2010.

[Bi] A. Bikjalis. An estimate for the remainder term in the central limit theorem. (Russian) Litovsk. Mat. Sbornik, 4 (1964), 303-308.

[Bl] N.M. Blachman. The convolution inequality for entropy powers. IEEE Trans. Inform. Theory, 11 (1965), 267-271.

[B-C-G] S.G. Bobkov, G.P. Chistyakov and F. Götze. Non-uniform bounds in local limittheorems in case of fractional moments. arXiv:1104.3759v1 [math.PR],(2011).

[B-G] S.G. Bobkov and F. Götze. Concentration inequalities and limit theorems for randomized sums. Probab. Theory Related Fields, 137 (2007), no. 1–2, 49-81.

[C-D-T] T.M. Cover, A. Dembo and J.A. Thomas. Information-theoretic inequalities. IEEE Trans. Inform. Theory, 37 (1991), no. 6, 1501–1518.

[E] C-G. Esseen. Fourier analysis of distribution functions. A mathematical study of the Laplace-Gaussian law. Acta Math., 77 (1945), 1-125.

[F] W. Feller. An introduction to probability theory and its applications. Vol. II. Second edition, John Wiley & Sons, Inc., New York-London-Sydney 1971, xxiv+669 pp.

[G-K] B.V. Gnedenko and A.N. Kolmogorov. Limit distributions for sums of independent random variables. Translated and annotated by K. L. Chung. With an Appendix by J. L. Doob. Addison-Wesley Publishing Company, Inc., Cambridge, Mass., 1954. ix+264 pp.

[G-H] F. Götze and C. Hipp. Asymptotic expansions in the central limit theorem under moment conditions. Z. Wahrscheinlichkeitstheorie und Verw. Gebiete, 42 (1978), no. 1, 67-87.

[I-L] I.A. Ibragimov and Ju.V. Linnik. Independent and stationarily connected variables. Izdat. "Nauka", Moscow, 1965, 524 pp.

[J] O. Johnson. Information theory and the central limit theorem. Imperial College Press, London, 2004, xiv+209 pp.

[L] Yu.V. Linnik. An information-theoretic proof of the central limit theorem with the Lindeberg condition. Theory Probab. Appl., 4 (1959), 288-299.

[M] V.K. Matskyavichyus. A lower bound for the rate of convergence in the central limit theorem. (Russian) Theory Probab. Appl., 28 (1983), no. 3, 565-569.

[O-P] L.V. Osipov and V.V. Petrov. On the estimation of the remainder term in the central limit theorem. (Russian) Theory Probab. Appl., 12 (1967), 322-329.

[Pe1] V.V. Petrov. On local limit theorems for sums of independent random variables. Theory Probab. Appl., 9 (1964), No.2, 312-320.

[Pe2] V.V. Petrov. Sums of independent random variables. Springer-Verlag, New York, Heidelberg, Berlin, 1975, x + 345 pp.

[Pr] Yu.V. Prohorov. A local theorem for densities. (Russian) Doklady Akad. Nauk SSSR (N.S.) 83, (1952), 797-800.

[S-M] S.H. Sirazhdinov and M. Mamatov. On mean convergence for densities. (Russian) Theory Probab. Appl., 7 (1962), No. 4, 433-437.

[Sz] G. Szegö. Orthogonal polynomials. Third edition. Amer. Math. Soc. Colloquium Publications, Vol. 23. Amer. Math. Soc., Providence, R.I., 1967, xiii+423 pp.

[T] H.G. Tucker. On a necessary and sufficient condition that an infinitely divisible distribution be absolutely continuous. Trans. Amer. Math. Soc. 118 (1965), 316-330.

Sergey G. Bobkov
School of Mathematics, University of Minnesota
127 Vincent Hall, 206 Church St. S.E., Minneapolis, MN 55455 USA
E-mail address: `bobkov@math.umn.edu`

Gennady P. Chistyakov
Fakultät für Mathematik, Universität Bielefeld
Postfach 100131, 33501 Bielefeld, Germany
E-mail address: `chistyak@math.uni-bielefeld.de`

Friedrich Götze
Fakultät für Mathematik, Universität Bielefeld
Postfach 100131, 33501 Bielefeld, Germany
E-mail address: `goetze@mathematik.uni-bielefeld.de`